

Universidade do Minho
Escola de Letras,Artes e Ciências Humanas

Diana Cristina Almeida Oliveira Neiva

Film as philosophy: philosophizing through cinematic thought experiments

junho de 2026



Film as philosophy: philosophizing through cinematic thought experiments

Diana Cristina Almeida Oliveira Neiva

UMinho | 2026

Universidade do Minho
Escola de Letras,Artes e Ciências Humanas

Diana Cristina Almeida Oliveira Neiva

Film as philosophy: philosophizing through cinematic thought experiments

Tese de Doutoramento
Doutoramento em Filosofia

Trabalho efetuado sob a orientação do
Professor Doutor Vítor Manuel Ferreira Ribeiro de Moura
da
Professora Doutora Susana Isabel Rainho Viegas
e do
Doutor Thomas William McClelland

junho de 2026

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.



CC BY-NC-ND

https://creativecommons.org/licenses/by-nc-nd/4.0/

Acknowledgements

First and foremost, I extend my deepest gratitude to the Foundation for Science and Technology (FCT) for the funding without which this project would not have been possible.

To my supervisors: Vítor Moura, for your tireless help, constant feedback, assistance with concepts and arguments, and even bureaucratic support. Susana Viegas, for your endless inspiration and encouraging conversations, an infinite source of knowledge in film and philosophy, and for the opportunities you've provided me to engage with the field. Tom McClelland, for your precious feedback, motivation, and even the help during my initial application stage, particularly during my Erasmus+ research period at the University of Warwick, where you encouraged me to pursue this journey. To all three of you, for your irreplaceable expertise, constant presence, the confidence you instilled in me, and your friendship.

To Thomas Wartenberg, for your inspiration, encouragement, and conversations that always sparked new ideas in my thinking. Once again, for your friendship.

To my family, especially my parents and grandparents, who have always motivated, believed in, and helped me in all my "adventures", including this one. To my husband, Adriano, who has been my rock, a constant and serene support that balanced me and allowed me to complete this project. Without him, I would not have found such calm waters.

To my friends, especially Raquel, another rock who motivates, inspires, and supports me unconditionally. To Daniela and Chris, who are always there for everything and keep my lighter side alive, something I desperately need to stay "sane". To Guida, who has been a continuous presence for so many years, always believing in me and inspiring me.

And of course, to my cats, who have been constant companions and have likely become experts in film as philosophy by now, given the time they've spent beside me while I wrote this.

´

Agradecimentos

Em primeiro lugar, expresso o meu mais profundo agradecimento à Fundação para a Ciência e Tecnologia (FCT) pelo financiamento sem o qual este projeto não teria sido possível.

Aos meus orientadores: Vítor Moura, pelo seu auxílio incansável, feedback constante, ajuda com conceitos e argumentos e até com o apoio burocrático. Susana Viegas, pela inspiração

sem limites e conversas encorajadoras, uma fonte infinita de conhecimento em cinema e filosofia, e pelas oportunidades que me proporcionou para contactar com a área. Tom McClelland, pelo feedback precioso, pela motivação e até pela ajuda na fase inicial da candidatura, particularmente durante o meu período de investigação Erasmus+ na Universidade de Warwick, onde me encorajou a seguir esta jornada. A todos os três, pela vossa expertise insubstituível, presença constante, confiança que me transmitiram, e pela vossa amizade.

A Thomas Wartenberg, pela inspiração, encorajamento e conversas que sempre fizeram despertar novas ideias no meu pensamento. Mais uma vez, pela amizade.

À minha família, especialmente aos meus pais e avós, que sempre me motivaram, acreditaram e ajudaram em todas as minhas "aventuras", incluindo esta. Ao meu marido, Adriano, que tem sido o meu pilar, um apoio constante e sereno que me equilibrou e permitiu concluir este projeto. Sem ele, não teria encontrado águas tão calmas.

Aos meus amigos, especialmente à Raquel, outro pilar que me motiva, inspira e apoia incondicionalmente. À Daniela e ao Chris, que estão sempre presentes para tudo e mantêm o meu lado mais leve, algo de que desesperadamente preciso para me manter "sã". À Guida, que tem sido uma presença constante ao longo de tantos anos, sempre acreditando em mim e a ser uma inspiração

E, claro, aos meus gatos, que têm sido companheiros constantes e que já devem ser especialistas em cinema como filosofia, dado o tempo que passaram ao meu lado enquanto eu escrevia isto.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

DECLARATION ON THE USE OF ARTIFICIAL INTELLIGENCE (AI) TOOLS

I hereby declare having acted this academic work with integrity and confirm that I have not resorted to any form of undue use of AI tools, nor to the falsification or omission of information, sources, data, results or processes in any of the stages leading to its elaboration.

I further declare that any use of AI tools was done in a transparent and responsible manner, being duly noted in the work, and that I critically verified the accuracy of the content produced with their support, assuming full responsibility for what I submit.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho, as well as the applicable standards of academic integrity and data protection.

Film as philosophy: philosophising through cinematic thought experiments

Abstract

This project is dedicated to the concept of film as philosophy, arguing that film can serve as a legitimate medium for doing philosophy. Through the notion of the cinematic thought experiment, it seeks to defend a revised and moderately bold thesis, showing that certain films can produce philosophical thought autonomously. Building on the tradition of Cavell and Deleuze, and engaging with the contemporary *film as philosophy* (FAP) debate, the research analyses the main sceptical objections — namely those of generality, rationality, and interpretive imposition — and proposes that these can be addressed through the mental modelling theory of thought experiments.

The investigation examines how films, through their exploration of time, narrative, and emotion, can offer unique conditions for philosophical reflection, complementing and sometimes extending traditional methods of textual argumentation. Drawing on case studies, particularly various films by Wes Craven, it aims to demonstrate that film can enact visual and narrative thought experiments that promote new forms of philosophical creation.

Keywords: film as philosophy; thought experiments; mental modelling; bold thesis; Wes Craven

Cinema como filosofia: filosofar através de experiências mentais cinematográficas

Resumo

Este projeto dedica-se ao conceito de cinema como filosofia, defendendo que o cinema pode constituir um meio legítimo de fazer filosofia. Através da noção de experiência mental cinematográfica, procura-se sustentar uma versão revista e moderadamente audaciosa, mostrando que determinados filmes podem produzir pensamento filosófico de forma autónoma. Partindo da tradição de Cavell e Deleuze, e do debate contemporâneo em torno do *film as philosophy* (FAP), analisam-se as principais objeções cépticas — nomeadamente as da generalidade, racionalidade e imposição interpretativa — e propõe-se que estas podem ser respondidas através da teoria dos modelos mentais das experiências mentais.

A investigação examina de que modo os filmes, ao explorarem tempo, narrativa e emoção, podem oferecer condições únicas para a reflexão filosófica, complementando e, por vezes, estendendo os métodos tradicionais de argumentação textual. Com recurso a estudos de caso, particularmente diversos filmes de Wes Craven, pretende-se demonstrar que o cinema pode realizar experiências mentais visuais e narrativas que promovem novas formas de criação filosófica.

Palavras-chave: cinema como filosofia; experiências mentais; modelos mentais; tese audaciosa; Wes Craven

Table of contents

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS iv

Acknowledgements v

STATEMENT OF INTEGRITY vii

Abstract viii

Resumo ix

Introduction 1

The film as philosophy debate 2

The appeal to thought experiments 5

Central thesis and arguments 6

Methodology and interpretive approach 7

Chapter overview and objectives 8

Implications 9

1. What is film as philosophy? 11

1.1 The film-philosophy idea 11

1.2 Sinnerbrink's approach 19

1.3 Revising the dichotomy 20

1.4 Mulhall's case for films as philosophical works 21

1.5 Wartenberg's approach to film as philosophy 24

1.6 Film as philosophical illustration 28

1.7 Structural film and self-reflection 32

2. Preliminary objections against film as philosophy and the appeal to thought experiments 40

2.1 The banality objection 40

2.2 The explicitness objection 44

2.3 The generality objection 50

2.4 The appeal to thought experiments 52

2.5 The role of documentary in film as philosophy 58

3. Thought experiments in science and philosophy 62

3.1 The paradox of thought experiments: the Platonic versus the Argument view 63

3.2 Middle ground positions, or moderate inflationism 74

3.3 Scepticism against thought experiments, or extreme and moderate deflationism 83
3.4 The role of intuition in thought experiments 97
3.5 Types of thought experiments 103
3.5.1 Counterexamples to philosophical theses 104
3.5.2 Establishing possibility 105
3.5.3 Demonstrating impossibility 106
3.5.5 Theory confirmation 107
3.5.6 Theory construction 109
4. Artistic thought experiments and the "bold thesis" 112
4.1 Artistic thought experiments 112
4.2 The rationality objection and the "bold thesis" 119
4.2.1 The rationality objection 119
4.2.2 Bold thesis: the means condition and the results condition 121
4.2.3 Dilemma of the paraphrase 123
4.2.4 The means condition and medium specificity 124
4.2.5 The results condition 128
4.2.6 Solving the dilemma 129
4.2.7 A revised bold thesis 131
4.3 Cinematic philosophy through thought experimentation 133
4.3.1 The value of thought experiments in philosophy 135
4.3.2 The problem of ambiguity, and the "Socratic model" 136
4.3.3 The issue of detail 139
4.3.4 The role of emotion and the "proper domains" of psychological capacities 141
4.3.5 Mental modelling through film 146
4.4 An integrated account of cinematic-philosophical thought experiments 150
5. The imposition objection and authorial intentions 154
5.1 Wartenberg's response: creator-oriented interpretation 158
5.2 Livingston's critique of Wartenberg 159
5.3 Types of intentionalism 161
5.3.1 Conditional intentionalism 161
5.3.2 Hypothetical intentionalism 162
5.3.3 Actual intentionalism and partial intentionalism 163
5.4 Livingston's interpretation of Bergman's work 165
5.5 Implications for film as philosophy 171

6. Thought experiments on screen: case studies **177**
6.1 Methodology and case selection 177
6.2 *The Last House on the Left*: a counterexample to common moral intuitions 182
6.2.1 Targeting revenge theory 183
6.2.2 The film's counterexample structure 184
6.2.3 Cinematic techniques 186
6.2.4 Collaborative reasoning 191
6.3 *A Nightmare on Elm Street*: establishing the possibility of a perceptual drift 198
6.3.1 Cinematic techniques 201
6.3.2 Collaborative reasoning 207
6.3.3 Cinematic advantages for possibility-establishing 208
6.4 *Scream*: metahorror, hyperreality and media effects 210
6.4.1 *Scream* as metahorror 212
6.4.2 *Scream* and the hyperreality possibility 218
6.4.2.1 The progressive collapse of reality: Scream (1996) 219
6.4.2.2 The progressive collapse of reality: Scream 2 (1997) 221
6.4.2.3 The progressive collapse of reality: Scream 3 (2000) 222
6.4.2.4 The progressive collapse of reality: Scream 4 (2011) 224
6.4.3 *Scream* as a counterexample to media effects assumptions 226
6.4.5 Cinematic techniques in *Scream*'s thought experiments 231
6.4.5 Collaborative Reasoning in *Scream* 235
6.5 *Wes Craven's New Nightmare*: an archetypal media theory 236
6.5.1 Evil archetypes and the making of meaning 238
6.5.2 Censorship and authorial responsibility 244
6.5.3 Philosophical synthesis and legacy 246
6.6 Summary and final considerations 247
7. Metaphilosophical considerations **254**
7.1 The problems of essentialist metaphilosophies 255
7.1.1 Restrictive metaphilosophies 257
7.1.2 Inclusive metaphilosophies 260
7.2 Alternatives to essentialist metaphilosophies 262
7.3 A moderate metaphilosophical approach 265
7.3.1 Moderate pro-film as philosophy 270
Conclusion **276**

References **283**

Film and Philosophy Literature 283

Thought Experiments 287

Scientific thought experiments 290

Aesthetics and Cognitive Value of Art 291

General Philosophy 293

Literature on Wes Craven and his films 295

Filmography 297

Figure sources 298

Introduction

> "I think films are dreams. They're manufactured realities that we created to help allay our fears and deal with our terrors in a magical way. I find that the mixture of dreams within films does precisely that." (Craven 1988).

The idea that films philosophise began, for me, in an unlikely place. I had always been drawn to the intersections between philosophy and popular culture, wondering whether the philosophical insights I found in zombie apocalypses and horror films, in general, were really there or whether I was simply reading too much into my favourite kind of entertainment. This question took on new interest about ten years ago during a class with Vítor Moura on art, representation and meaning, where I found myself arguing that *The Walking Dead* could be understood as a serious philosophical TV series. Rather than dismissing this as student enthusiasm, Moura suggested I explore Thomas Wartenberg's *Thinking on Screen: Film as Philosophy* (2007), where the idea that films themselves might philosophise rather than merely providing material for philosophical reflection was defended.

That conversation fundamentally shifted my approach to both philosophy and film. What started as a general intuition that films might be philosophical developed into a more precise question: can films actually *do* philosophy in ways that complement or even extend traditional philosophical methods? This thesis argues that they can, specifically through their capacity to function as philosophical thought experiments.

The question of whether films can philosophise has generated considerable debate within philosophy of film over the past few decades. Defenders of what has become known as the "film as philosophy" (FAP) position argue that certain films engage in genuine philosophical thinking. Sceptics counter that films lack the argumentative rigour necessary for philosophical work and that any philosophical content we find in them is imposed by interpreters rather than present in the films themselves.

This debate intersects with broader questions about the nature of philosophical methodology that are relevant to our present day. As visual and narrative media become increasingly central to how ideas are developed and transmitted, the question of whether philosophical thinking can occur through cinematic means becomes pressing for understanding both the future of philosophy and the intellectual resources available within popular culture.

The works of Stanley Cavell and Gilles Deleuze influenced the foundations for these questions. Cavell's engagement with cinema in *The World Viewed* (1971) and his studies of Hollywood genres suggested that films actively participate in philosophical reflection about the human condition. Deleuze's *Cinema* volumes offered a different approach, treating films as genuine forms of thought that create concepts through their distinctive manipulation of movement and time. These approaches also established different ways of understanding film's philosophical potential.

By the turn of the century, the journal *Film-Philosophy* gave these ideas new momentum, and the question took new directions, focusing on "To what extent can film – or individual films – act as a vehicle of or forum for philosophy itself?" (Wartenberg and Smith 2006, 1).

The film as philosophy debate

The contemporary FAP debate emerged from both an increasing dissatisfaction with traditional "philosophy of film" approaches that treated cinema primarily as an object of philosophical analysis rather than as a potential source of philosophical insight, and an interest in the idea that films may have their own philosophical voice. There are several degrees of commitment to a FAP thesis. One may simply enjoy watching films or television series and view them as inspiring philosophical debates, as exemplified in the *Popular Culture and Philosophy* series published by Open Court Publishing Company, with books such as *The Good Place and Philosophy* (Benko and Pavelich 2019) or *Blade Runner 2049 and Philosophy* (Bunce and McCrossin 2019). More ambitious versions not only relate film to philosophy but also endorse that some films are philosophy themselves, meaning that films may be works of philosophy, that filmmakers philosophise through their films, that all cinema is inherently philosophical by nature, or that cinema thinks philosophically in its distinctive way.

The challenge to the hierarchy that poses philosophy above film as an intellectual endeavour has taken several forms. Robert Sinnerbrink's "film-philosophy" approach emphasises cinema's capacity for "distinctively cinematic thinking" that transforms philosophical concepts. While Sinnerbrink's work offers valuable insights into cinema's transformative potential, his dichotomy that opposes "philosophy of film" to "film-philosophy" may exclude important contributions from analytic and cognitivist philosophy of film, which have also developed approaches to understanding cinema's cognitive and philosophical value.

Stephen Mulhall's case-by-case analyses of specific films provide a different model for understanding cinematic philosophy. His reading of the *Alien* tetralogy as a philosophical investigation of human personal identity and embodiment in *On Film* (2002/2008) demonstrates how films can engage philosophical questions through their distinctive narrative and visual techniques without requiring explicit philosophical argumentation. Mulhall's approach suggests that films can philosophise through their particular ways of presenting and exploring concepts.

Agreeing with this case-study approach, Thomas Wartenberg has developed one of the most systematic investigations into film as philosophy, focusing on specific techniques that filmmakers can employ to philosophise. He makes a distinction between "film as philosophy" and "philosophy in film", where the former claims that films themselves can philosophise, and the latter might simply mean philosophy is present in a film (such as characters engaging in philosophical dialogue). Wartenberg's moderate pro-film as philosophy approach avoids sweeping claims about cinema's inherent philosophical capacity, instead examining how particular films employ cinematic techniques to contribute to philosophical understanding.

However, the FAP thesis faces substantial objections. The banality objection, articulated by Jerome Stolnitz in *On the Cognitive Triviality of Art* (1992), contends that works of fiction, including films, typically present familiar truths but fail to produce any philosophical knowledge. According to this view, what we take to be philosophical content in films is usually commonplace trivialities dressed up in a compelling narrative.

The explicitness objection, advanced by Murray Smith in "Film Art, Argument, and Ambiguity" (2006) and Bruce Russell in "The Philosophical Limits of Film" (2008), argues that films cannot construct the explicit arguments that philosophical thinking requires. This objection maintains that philosophical work must involve a clear articulation of explicit arguments, something that narrative films typically do not provide. Films may raise philosophical questions or present philosophical scenarios, but they cannot engage in the rigorous argumentation that identifies philosophical thinking.

The generality objection maintains that films present particular scenarios that cannot support the universal claims characteristic of philosophical theorising. Bruce Russell has argued that films' focus on specific characters and situations prevents them from establishing the general principles that philosophical theories require. Even when films present scenarios that seem philosophically relevant, whether because they raise philosophical questions or may constitute a

counterexample to a philosophical thesis, their particularistic nature and lack of explicit argumentation limit their capacity to contribute to philosophical understanding.

The imposition objection accused FAP enthusiasts of imposing their philosophical interpretations onto the films. This objection strikes at the heart of the FAP hypothesis by suggesting that the philosophical work is being done by philosophers who use films as facilitators for their own thinking rather than by the films themselves. According to this view, films provide raw material for philosophical reflection but do not themselves engage in philosophical thinking.

Another interesting challenge is the rationality objection and its related "bold thesis" proposed by Paisley Livingston in "Theses on Cinema as Philosophy" (2006). We may ask why we would watch a film for philosophical purposes instead of just reading a traditional philosophical work. Thus, Livingston proposed two conditions for a film to philosophise: films must make a philosophical contribution through exclusive capacities of the cinematic medium (the means condition). That contribution must be innovative and independent (the results condition).

These objections collectively challenge film's capacity to engage in genuine philosophical work, setting the stage for examining how the appeal to thought experiments might address these concerns.

These debates intersect with several enduring philosophical discussions about the cognitive value of art that have been debated since Plato's critique of poetry in the *Republic*. The traditional philosophical suspicion of art's epistemic power continues to influence contemporary discussions about whether artistic works can contribute to philosophical understanding or merely provide entertainment and emotional experience.

Philosophers like Peter Lamarque and Stein Haugom Olsen (1994) have defended anti-cognitivist positions that treat artistic and cognitive value as fundamentally distinct, arguing that appreciating art as art requires setting aside concerns about truth and knowledge. Conversely, philosophers like Noël Carroll (2002), Catherine Elgin (1993), and Eileen John (1998) have defended various forms of artistic cognitivism, arguing that artworks can contribute to human understanding through their distinctive modes of presentation.

The film as philosophy debate also connects to broader questions about philosophical methodology and the own nature and purpose of philosophy: if films can philosophise, this challenges assumptions about what counts as philosophical thinking and raises questions about philosophy's methods.

The appeal to thought experiments

My approach to these challenges focuses on the appeal to thought experiments as a promising response to FAP scepticism. Against the various objections to cinematic philosophy, proponents like Thomas Wartenberg (2007), Noël Carroll (2006), and Tom McClelland (2019) argue that some films can be seen as philosophical thought experiments, using examples such as *Serene Velocity* (Ernie Gehr, 1970) *and The Matrix (Wachowski Sisters,* 1999). This appeal draws parallels between cinematic narratives and the thought experiments that play central roles in philosophy.

Thought experiments are "means of imaginative reasoning" (Stuart et al. 2018) that have played a central role in philosophical and scientific inquiry. The epistemological puzzle they present – how purely conceptual exercises can yield knowledge about the world – has generated extensive debate about their nature and value. The debate is currently marked by two main opposing views: Platonic (Brown 2004) and empiricist (Norton 2004). The Platonic view treats thought experiments as providing direct access to laws of nature through rational intuition. In contrast, the empiricist view argues that thought experiments are merely disguised arguments that derive conclusions from premises already accepted by their readers.

Between these extremes, naturalistic views offer more moderate accounts of thought experiments' epistemic value. The mental modelling account developed by Nenad Miščević (1992) and Nancy Nersessian (2018) emphasises how imaginative scenarios enable us to manipulate and reorganise existing knowledge structures. Tamar Gendler's (2000) constructivist approach argues that thought experiments generate "quasi-observational" knowledge by enabling us to track features of imaginary scenarios. According to David Davies (2018), such naturalistic accounts are compatible with a moderate inflationist view of thought experiments, which offers the best method to defend the claim that thought experiments can provide knowledge despite being fictional narratives.

However, appealing to thought experiments also raises complex questions, as traditional philosophical thought experiments have themselves been subject to various criticisms. Kathleen Wilkes has argued that philosophical thought experiments often rely on inadequately specified background conditions that make their conclusions unreliable (1988), while Daniel Dennett has criticised the philosophical reliance on "intuition pumps" that manipulate our responses through selective presentation rather than genuine reasoning (1995).

The thought experimental approach offers several advantages over previous defences of cinematic philosophy. It positions films as tools for philosophical investigation that can generate

new insights through their cinematic techniques. Cinematic thought experiments may offer solutions to the traditional limitations of thought experiments through their rich audiovisual presentation, detailed scenario construction, and engagement with emotional and embodied understanding.

Central thesis and arguments

This thesis develops and defends a revised bold thesis about cinematic philosophy: films can produce philosophy through thought experimentation, and they do so in ways that are characteristically cinematic rather than identical to traditional philosophical practice. Furthermore, they do not replace, but complement traditional methods of philosophising. My argument advances through several interconnected claims that challenge both strong sceptical positions about cinema's philosophical potential and overly optimistic claims about cinema's inherent philosophical nature.

I contend that many traditional objections to film as philosophy rest on overly restrictive conceptions of philosophical method that privilege explicit argumentation over other forms of philosophical investigation. While explicit argumentation plays an important role in philosophical practice, it is not the only legitimate form of philosophical thinking. Thought experiments, conceptual analysis in general, and other methods of philosophical investigation operate through different modes of presentation and engagement while still contributing to philosophical understanding.

I also argue that films can function as philosophical thought experiments when they employ the methodological structure of thought experimentation, address philosophical subjects, and generate philosophical understanding through the collaborative interaction between cinematic techniques and viewer interpretation. This requires careful analysis of how specific films use cinematic techniques to construct scenarios that test philosophical concepts and generate philosophical insights.

Against strong sceptical positions that deny film's philosophical potential a priori, I maintain that questions about specific films' philosophical status should be addressed through examination of particular works rather than theoretical arguments about cinema's general capacities. Against strong pro-film positions that claim cinema in general is inherently philosophical, I maintain that whether a film philosophises depends on careful analysis of how the film employs cinematic techniques to address philosophical questions.

This moderate approach includes several key commitments: a pluralistic essentialist account of philosophy that recognises multiple legitimate forms of philosophical practice while maintaining that not all forms of thinking qualify as philosophical; a preference for particularist approaches that focus on case-by-case analyses of particular films rather than sweeping generalisations about cinema's philosophical capacities; and a complementary thesis that regards different media as capable of making distinctive but complementary contributions to philosophical understanding.

Methodology and interpretive approach

My methodology reflects this moderate approach through detailed analyses of specific films that demonstrate how cinematic techniques can generate philosophical understanding. Rather than offering sweeping generalisations about cinema's philosophical capacities, I focus on case studies that show how particular films employ thought-experimental methods to address philosophical questions.

The interpretive approach I develop integrates insights from moderate intentionalism with attention to cinema's distinctive techniques. Following Vítor Moura's analysis of cinematic communication (2021), I treat films as instances of "nonnatural meaning" that depend on recognisable intentions while operating through cinematic techniques and cultural contexts. Moura's concept of the "spectral author" – a presence that emerges from the work itself while retaining connection to the actual director – provides a basis for understanding how films create communicative experiences.

This approach avoids both extreme author-centred positions that treat films' meanings as entirely determined by filmmakers' explicit intentions and extreme reader-response approaches that make meaning entirely dependent on viewer projection. Instead, I develop a moderate intentionalist position that recognises filmmakers' role in establishing conditions for philosophical investigation through cinematic techniques while acknowledging viewers' active participation in developing philosophical implications.

Throughout the thesis, I maintain focus on cinema's characteristic techniques, such as editing, framing, temporal manipulation, sound design, and mise-en-scène, examining how these techniques contribute to the philosophical investigation in question.

Chapter overview and objectives

The thesis progresses through seven chapters, each building upon the theoretical foundations established in preceding sections.

Chapter 1 examines existing approaches to film as philosophy, providing a basis for subsequent arguments through analysis and critique of Sinnerbrink's film-philosophy approach, Mulhall's case-by-case method, and Wartenberg's moderate theory for film as philosophy, which focuses on specific techniques that filmmakers can employ to philosophise, most centrally, illustration and thought experimentation.

Chapter 2 addresses preliminary objections to film as philosophy and introduces the appeal to thought experiments as a promising response. I examine the banality objection, the explicitness objection, and the generality objection, exploring how the appeal to thought experiments offers a potential resolution to these concerns while raising new questions about the nature of thought experimental reasoning.

Chapter 3 explores thought experiments in science and philosophy, establishing the foundation for understanding how thought experiments function and what makes them valuable in philosophy. We examine the paradox of thought experiments and investigate competing explanations of their epistemic value, including debates between Platonic and argument views, and moderate inflationist positions that offer middle-ground explanations. This analysis helps establish criteria for evaluating thought experiments' effectiveness and identifies ways that cinematic presentation might address traditional limitations.

Chapter 4 develops a core argument about artistic thought experiments and presents our revised bold thesis. We examine how artistic works can function as philosophical thought experiments and address the rationality objection while examining and revising the bold thesis. Through mental modelling accounts, I demonstrate how films can activate tacit knowledge and reorganise conceptual understanding through quasi-observational processes, establishing necessary conditions for films to function philosophically.

Chapter 5 tackles the imposition objection through an exploration of authorial intentions and interpretation. I develop a moderate intentionalist approach that enables collaborative reasoning between filmmakers and viewers while avoiding extremes of both author-centred interpretation and reader-response approaches. This interpretive method addresses the imposition objection by grounding interpretation in evidence from films themselves while acknowledging the collaborative nature of meaning-making in cinema.

Chapter 6 applies established theoretical grounds through detailed case studies, focusing primarily on Wes Craven's horror films to demonstrate how specific films function as different types of philosophical thought experiments. I examine *The Last House on the Left* (1972), *A Nightmare on Elm Street* (1984), *Scream* (1996), and *Wes Craven's New Nightmare* (1994), showing how each film engages with distinct philosophical questions through characteristically cinematic means.

Chapter 7 explores the metaphilosophical foundations underlying the film as philosophy debate, arguing that the controversy cannot be resolved by examining cinema's capacities alone but also requires confronting fundamental questions about the nature of philosophy itself. I defend a moderate metaphilosophical position that maintains philosophical standards while remaining open to innovation.

Implications

The relevance of this investigation extends beyond academic debates about film and philosophy to encompass fundamental questions about the nature of philosophy itself and the recognition of intellectual work occurring outside traditional academic contexts. If films can philosophise through thought experimentation, this challenges traditional assumptions about philosophical methodology and reveals substantial philosophical resources within contemporary culture that academic philosophy has largely overlooked.

The methodological challenge has particular implications for how we understand the relationship between popular culture and serious intellectual discourse. Academic philosophy has historically maintained boundaries between rigorous philosophical work and popular forms of cultural expression, treating the latter as suitable subjects for philosophical analysis but not as potential sources of philosophical insight themselves. If films can function as philosophical thought experiments, these boundaries may need reconsidering, opening new possibilities for recognising intellectual work that occurs outside traditional academic contexts.

The urgency of these questions is amplified by contemporary culture's increasing reliance on visual and narrative media. The rise of streaming platforms, virtual reality technologies, and interactive media suggests that visual storytelling will play an increasingly central role in how ideas are developed and transmitted. If philosophical thinking can occur through cinematic means, then we need theoretical approaches capable of recognising and evaluating such thinking as these new media forms continue to evolve.

Beyond these cultural considerations, the investigation also addresses longstanding concerns about thought experiments themselves. Through recognition of cinema's philosophical potential, we can identify substantial philosophical work that has been occurring within forms of popular culture typically dismissed by academic philosophy. This reveals new resources for addressing longstanding philosophical questions about consciousness, reality, moral psychology, and human nature while demonstrating that serious philosophical work can occur in the most unexpected places, even in the horror films that first sparked my interest in the intersections between philosophy and popular culture.

Through this contribution, I hope to encourage broader recognition of the intellectual resources available within contemporary culture and to support the ongoing evolution of philosophical practice in response to changing cultural and technological conditions.

1. What is film as philosophy?

1.1 The film-philosophy idea

In *New Philosophies of Film: Thinking Images* (2011), Robert Sinnerbrink examined the subject of film as philosophy as part of a more extensive discussion about how film and philosophy should relate in general. He proposed a new perspective on the relationship between philosophy and film that challenges the commonly accepted "Film-Philosophy" approach. This new approach avoids creating a division between the "Continental" and "Analytical" approaches to this relationship. Rather than privileging either of those approaches, film-philosophy emphasises that philosophical claims about film cannot be decided purely on theoretical grounds, but require aesthetic analysis of films. This new approach focuses on how films themselves think, rather than applying pre-existing philosophical theories. He argues that film-philosophy resists "reducing cinema to a mere reflection of a philosophical thesis or framework" (Sinnerbrink 2011, 137), allowing instead films to challenge philosophical theories through their cinematic thinking.

Other advocates of this approach share the same motivation of avoiding a dichotomy between continental and analytical approaches. David Sorfa, for instance, despite admitting that many of the researchers in this approach arrived at Film-Philosophy "through structuralism, post-structuralism and the liberationist critiques promised by Michel Foucault, Jacques Lacan, Jacques Derrida, Gilles Deleuze and Félix Guattari" (Sorfa 2016, 4), also denies this dichotomy. David Martin-Jones, too, argues that the division between Continental Philosophy and Analytic Philosophy is erroneous and Eurocentric (Martin-Jones 2016, 8).

Despite claiming that a dichotomy between "continental" and "analytical" is avoided, Sinnerbrink sets up an opposition between what he calls the "Culturalist-Historicist" approach *versus* the "Cognitivist-Naturalist" approach. The first approach would be one that Noël Carroll and David Bordwell called the "Grand Theory", and the second approach initiated by the same philosophers is one that Sinnerbrink sometimes names Analytic-Cognitivist (Sinnerbrink 2011, 2), sometimes Cognitivist-Naturalist (opposed to the Culturalist-Historicism) or Anglo-American (Ibid., 5).

While the Culturalist-Historicist approach consists of "cultural studies, media theory, post-colonialism, gender studies, queer theory, reception studies, production histories, historical and cross-cultural approaches, film history, especially early cinema, and so on" (Ibid., 4), the

"Naturalist" approach consists of "'post-Theory', cognitivism, analytic philosophy and, more recently, neuroscience and evolutionary biology [which] have begun to coalesce into a formidable research paradigm. The latter maintains that film studies, ideally, should draw on the best available science, be compatible with philosophical naturalism, and demonstrate cumulative, testable results." (Ibid., 4)

Philosophy of Film in the Analytic-Cognitivist tradition, founded by Noël Carroll and David Bordwell, has emerged as an alternative to the "Grand Theory" of a continental base that emerged in the seventies "centrally composed of elements of Louis Althusser, Jacques Lacan, and Roland Barthes, often with optional features derived, often incongruously, from Michel Foucault, Julia Kristeva, Pierre Bourdieu, Gilles Deleuze [etc.]" (Bordwell and Carroll 1996, 37).

The alternative to this, according to Carroll, is known as cognitivism. Cognitivism, as Carroll describes, prioritises conscious mental mechanisms in explaining how viewers interpret and comprehend films, in contrast to approaches that rely on unconscious psychological processes (Carroll 1996, 321). What this American philosopher seeks to highlight in this approach is the practice of "a mode of rational enquiry (...) according to our best standards of reasoning and evidence [being that] the theories are evaluated comparatively" (Ibid., 321). So this means that the various cognitivist film theories are not a conglomerate that composes one general theory that can be instantiated in films. It is, instead, a model that relies on empirical evidence and strong argumentation to approach issues on film. Thus, not only do the cognitivist theories compete and compare with the psychoanalytical ones, but they also compete with each other.

For example, to address the *paradox of fiction* – which questions why audiences can feel emotional responses to fictional works – there are various explanations within the cognitivist model. These explanations are considered cognitivist because they all emphasise the role of mental representations, rational thought processes, and cognitive judgments in our emotional responses to fiction. One influential idea is Kendall Walton's notion that "we do not just observe fictional worlds from without. We live in them (...)" (Walton 1990, 273). This suggests that we emotionally identify with the characters. This *identification theory* is further developed by Berys Gaut, who argues that we engage with films by imagining ourselves in the characters' situations. This explains why we are moved by their experiences despite knowing they are fictional (Gaut 1999). Similarly, Gregory Currie's view posits that emotional responses to fictional works are not genuine emotions in the traditional sense but rather "quasi-emotions" (Currie 183-84) which arise from a complex interplay of cognitive processes involving make-believe instead of belief. He argues that while

emotions typically require belief, readers engage with fiction through imaginative play that generates emotional-like responses: "Our responses to fiction differ from emotions in that make-belief is substituted for belief" (Ibid., 196). Currie emphasises that these quasi-emotions are analogous to genuine emotions as they involve feelings elicited by fictional events while being grounded in the reader's imaginative engagement, noting, "Quasi-emotions do not involve belief but are, in other ways, very like the emotions" (Ibid., 184). In essence, the reader's engagement with fiction allows for emotional experiences that mirror genuine emotions, albeit through a framework of make-believe rather than direct belief in the story's reality. So, Walton's cognitivist theory views our engagement as a rational game of make-believe requiring cognitive participation. Gaut's identification theory is cognitivist by emphasising the deliberate imaginative process through which we project ourselves into characters' situations. Currie's quasi-emotion theory is cognitivist, distinguishing between cognitive structures (belief versus make-belief) while maintaining that our emotional responses involve rational mental processes rather than mere reflexive reactions.

However, Noël Carroll challenges the identification hypothesis, arguing that our sympathy for fictional characters does not require identification. According to Carroll, audiences remain distinct from the characters because they are not in the same situation and often possess knowledge that the characters lack. Instead, the director "pre-focuses" or guides emotions through soundtracks, narrative structures, camera angles, and other cinematic techniques (Carroll 1999a). Carroll rejects the simulation theory, which suggests that audiences emotionally connect with characters by replicating or mirroring their emotions and intentions. Instead, he argues that our emotional responses to fiction are better understood through sympathy. Sympathy, for Carroll, involves a caring and concerned attitude toward fictional characters, characterised by emotional support and moral endorsement rather than direct emotional sharing; as he explains, "a matter of care, concern, or sympathy – an emotional state that we have for and not with the characters in question" (Carroll 2013, 50–51).

What Sinnerbrink points out about this kind of cognitivist approach is that it tends to disenfranchise film from any relevant contribution to philosophy. The so-called "*philosophical disenfranchisement of film*" would be an extension of "Philosophy's inveterate tendency to subordinate art as an inferior way of knowing, one that is theoretically completed by philosophy proper" (Sinnerbrink 2011, 128). This Analytic-Cognitivist tradition concerning the relationship between Film and Philosophy, according to Sinnerbrink, discusses this relationship from the perspective of a superior epistemic authority of philosophy before film (and, in general, art).

Sinnerbrink borrows the concept of "disenfranchisement" from Arthur Danto's work on disenfranchisement of art (Danto 2005), applying it specifically to film theory. Danto uses "disenfranchisement" to describe the philosophical tendency to strip art of its autonomous power, and, in this case, reduce artistic works, specifically films' philosophical potential, rather than recognising them as independent modes of thinking.

For Sinnerbrink, this disenfranchisement operates in many ways. First, the analytic-cognitivist approach tends to position film as merely an object of philosophical analysis rather than a potential source of philosophical insight in its own right. As he argues, "the attempt to subsume works of art into a philosophical discourse that enables us to master, comprehend and subordinate the work to theoretical or moral concerns" (Sinnerbrink 2001, 4-5) is precisely what characterises the disenfranchisement. The disenfranchisement is further evidenced in how analytic-cognitivist theories emphasise rational comprehension, tending to undervalue the aesthetic dimensions if film. He observes that there is a specific "charm to film which is not captured by 'rational reconstructions' nor 'cognitivist comprehension'" (Ibid., 53)

He suggests that instead we need "to rethink the hierarchical relationship between philosophy and art" and to explore how "our conventional understanding of philosophy [...] might be renewed and transformed (Ibid., 117-18).

Thus, the division that he seeks to preserve, instead of other more challenging divisions between Analytical and Continental, is a division between Philosophy of Film and Film-Philosophy, where the study of Philosophy of Film (such as formulation of a Philosophy of *x*) analyses, studies, reflects or theorises about film (Cf. Sinnerbrink 2011, 7-8), and Film-Philosophy regards film "as engaging in philosophically relevant reflection via the medium of film itself" (Ibid., 7). This kind of approach, for the philosopher, could be seen as a "rationalist" mode of theorising film: "Rationalist approaches to theory seek to provide explanatory models of various aspects of film experience. They elaborate empirically grounded models of our experience of moving images, of film ontology, of how we understand film narrative, and so on" (Ibid., 7-8).

So, instead of doing Philosophy of Film, Sinnerbrink advocates for doing Film-Philosophy that treats film as distinct from philosophy yet equally important and not subordinate to it. This, he argues, contrasts with the Philosophy of Film approach (cf. Ibid., 120). Sinnerbrink's approach is informed by the ideas of Stanley Cavell and Gilles Deleuze, whose shared focus is "why film matters philosophically" (Ibid., 90).

A significant thinker who addresses this is Siegfried Kracauer who believed that cinema plays a vital role in engaging with reality. For Kracauer, cinema helps restore our belief in the world's reality, which has been increasingly challenged by the disorienting effects of modern physics and scientific advancements (Kracauer 1997). Kracauer argues that film uniquely reveals the material reality of the world by capturing details that other art forms or philosophical approaches cannot. Film is grounded in perception and physical reality – which differs from Deleuze's view that cinema creates 'new dimensions of the real.' Where Deleuze sees cinema as generating previously inaccessible realities, Kracauer emphasises cinema's redemptive ability to reconnect us with the material world that already exists but has become obscured. For Kracauer, film does not primarily create new realities but provides a means of reconnecting with the world by showing it as it is. This is especially significant in an era dominated by scientific abstraction. Just as Cavell and Deleuze see film as a response to philosophical issues like scepticism and nihilism, Kracauer views cinema as a counter to the reduction of the world to abstract, mechanistic laws, helping us reconnect with the reality we experience.

Cavell views film as "an implicit response to scepticism that enacts a retrieval of the ordinary, thus providing an image for what philosophy strives to overcome" (Sinnerbrink 2011, 90). In Cavell's view, film helps to overcome scepticism – the philosophical problem of doubting the certainty of knowledge – which has dominated modern thought since the scientific revolution, secularism, and the Enlightenment (Ibid., 103). According to Cavell, film addresses scepticism by creating meaning for our reality, mainly through its engagement with the fictional "reality" it presents (Cavell 1971). Cavell argues that film allows us to engage with the ordinary world, offering a retrieval of the ordinary that gives significance to everyday experience. Rather than abstracting from reality, film captures the mundane and human experience, providing us with a renewed sense of certainty in the world as we experience it. This direct engagement with the ordinary counters philosophical scepticism by showing us a world we cannot directly inhabit, yet which reveals the truth of our everyday interactions, emotions, and perceptions. Film responds to scepticism not by offering absolute knowledge, but by portraying the ordinary from a perspective that simultaneously distances and connects us, revealing a meaningful world that we can trust despite philosophical doubts.

Like Cavell, who sees film as a means of addressing philosophical scepticism through its engagement with ordinary experience, Deleuze also recognises film's unique philosophical potential. However, where Cavell focuses on film's ability to retrieve and validate our everyday

experience of the world, Deleuze emphasises cinema's capacity to generate entirely new modes of thinking and perception. Despite their different approaches, both philosophers recognise film's power to address fundamental philosophical problems in ways that traditional philosophical writing cannot; Deleuze's conceptualisation of cinema, mainly through the movement-image and time-image (Deleuze 1986, 1989) distinction reflects his broader philosophical project of understanding how cinema can generate new modes of thinking and perception.

As Viegas notes, Deleuze sees cinema as capable of creating "a shock in thought by the communication of vibrations, directly reaching our nervous and cerebral system" (Viegas 2018, 280). The transition from movement-image to time-image cinema marks a philosophical shift. While the movement-image subordinates time to movement through logical, action-based sequences, the time-image presents time in its pure state, that is, a direct image of time no longer mediated by action or narrative (Deleuze 1989, 271). Deleuze conceptualises this temporality as non-chronological and split, where each movement divides into a present that passes and a past that is preserved. This is present in the "crystal-image": what we see in the crystal is time itself, the gushing forth of time (Ibid., 81). The crystal reveals the actual and virtual, present and past, as "distinct and yet indiscernible" (Ibid.) An example of this is Hitchcock's *Vertigo* (Ibid., 82) where the protagonist, Scottie, becomes trapped in a circuit of memory and repetition. His obsession with recreating the past, by transforming Judy into Madeleine, is an example of how past coexists with and distorts the present. In this case, the viewer is drawn into a non-chronological temporality, or a direct image of time.

This shift parallels broader changes in post-war philosophical thinking about subjectivity and reality. Importantly, Viegas highlights how Deleuze's theory moves beyond simple representation. The time-image does not merely represent different temporal experiences – it creates new ways of perceiving and thinking about time. As she explains, modern cinema achieves this by breaking with "the mechanistic structure and revealing the cliché qua cliché, thus creating a new cinematographic image that makes us think differently, that make us feel otherwise" (Viegas 2018, 279). For example, in Alain Resnais' *Hiroshima mon amour* (1959), the interplay between memory and the present disrupts the linearity of time. Here, the time-image does not just present a recollection of past events but makes us experience the past as something active and coexistent with the present. This non-chronological presentation of time invites us to reflect on the nature of memory and perception. We could also see this in Christopher Nolan's films. In *Interstellar*, for example, time is not experienced as a uniform, linear progression but as fractured and relative.

The scene on Miller's planet, where one hour equates to seven years on Earth, creates a visual and emotional disjunction. The characters' experiences of time are no longer mechanical or predictable but fluctuate in a non-chronological manner, forcing both the characters and the audience to confront time as something subjective and unpredictable. This representation of time shifts our perception as it is no longer a simple measurement but something that feels subjective, shaping the characters' relationships and decisions. As such, *Interstellar* illustrates Deleuze's point that modern cinema can present time in its pure, non-chronological state, making us rethink how we experience it.

This connects to Deleuze's response to nihilism:

> The cinema must film, not the world, but belief in this world, our only link. The nature of the cinematographic illusion has often been considered. Restoring our belief in the world - this is the power of modern cinema (when it stops being bad). Whether we are Christians or atheists, in our universal schizophrenia, we need reasons to believe in this world. It is a whole transformation of belief (Deleuze 1989, 172).

So, cinema, mainly through the time-image, can reactivate our connection to the world by creating what Viegas terms "non-human thoughts and non-human sensations" (Viegas 2018, 278). These new perceptual and cognitive experiences help overcome the passive nihilism of mere recognition and representation.

The idea is that film can create "non-human thought and sensations", or new ways of perceiving reality that do not rely on typical psychological frameworks. Again, *Interstellar* may be an example of a film that evoke non-human experience of time, and of existence in general. In the film, the representation of time dilation, where times flow depends on gravity, is a reconfiguration of how time is perceived. In the scenes where the main character, Cooper, experiences the passage of time, the audience is confronted with a reality that is abstract, not connected to human emotions. The experience of time in this film does not belong to a human subject, and his psychological experience does not bind it. On the contrary, time is presented as an external force, operating through a non-human mode of thinking. The sequence where Cooper watches decades pass in a matter of minutes, experiencing his children's ageing, reflects how the film itself thinks beyond human perception. This results in a new way of perceiving and thinking that goes beyond human

experience, enabling us to overcome nihilism by rethinking our relationship with time and with reality.

That films like these can create "non-human" thoughts and sensations, and that this enables us to reactivate our connection to the world, is arguable. We may argue that such experiences can actually create a sense of alienation or estrangement. The confrontation with abstract and impersonal temporalities may not reconnect or heal the rift between subject and world but intensify it. Thus, instead of offering a path beyond nihilism, films like *Interstellar* might reveal the limits of human understanding of and in the universe. In this sense, non-human thought may make a fracture more visible, aggravating some existential disorientation.

However, the significance of this approach lies in its suggestion that cinema can contribute to philosophical thinking in ways that traditional conceptual philosophy cannot. As Viegas argues, while films may not "philosophise" in the strict sense of creating concepts, they can create new "sensations that persist" and "affects and percepts that do not rely on the human" (Ibid., 272). This opens up new possibilities for addressing philosophical problems like nihilism through non-conceptual means. Deleuze's theory suggests that cinema is not simply a medium for representing thought, but one that *produces* thought. Rather than reconnecting us with already known reality, cinema helps us discover new dimensions of the real that were previously inaccessible to our experience. This function appears most evident in what Viegas identifies as cinema's ability to create "a possible world that is self-referential and autopoietic" (Ibid., 272). Through this, cinema does not merely depict reality, it makes it accessible to us in new ways, helping overcome what Viegas describes as the "intellectual passivity characteristic of the Cave prisoners" (Ibid.,279). In contrast to Kracauer and Cavell, who saw cinema as rescuing the real by reconnecting us with the world, Deleuze opens up new dimensions of reality that were previously inaccessible to our experience.

Thus, as Sinnerbrink notes, "[c]inema is philosophical, for both Deleuze and Cavell, because of the way it retrieves the ordinary and fosters the creation of new perspectives or modes of existence" (Ibid., 103). So, as Deleuze believes that film, as creating new concepts, can be philosophy, Cavell thinks that film has the same interest as philosophy in the problem of scepticism. In both cases, philosophy and film are viewed as mutually transformed in response to the problem of belief (Ibid., 20).

1.2 Sinnerbrink's approach

Defending this approach, Sinnerbrink retains that philosophy and film can transform each other, stating that film can be seen as engaging in philosophically relevant reflections through the medium or even in a "distinctively *cinematic* kind of thinking" (Ibid., 7). Film can then raise philosophical questions and force philosophy to reflect upon its nature. This happens in such a way that philosophy and film are regarded as existing at the same level (even if they are different) as involving a mutually transformative reflection:

> My suggestion is that the most productive way of exploring the idea of film as philosophy is an invitation to rethink the hierarchical relationship between philosophy and art. The encounter between film and philosophy invites us to explore novel ways in which our conventional understanding of philosophy (...) might be renewed and transformed (Ibid., 117-18).

For the philosopher, a film like Terrence Malick's *The New World* "cannot be reduced to any specific philosophical theme, idea or perspective." It is so "new" that it resists any translation into a known philosophical thesis. For Sinnerbrink, in this case, philosophy only "accompanies" film (Ibid., 181). Sinnerbrink, however, analyses and tries to interpret the new and untranslatable philosophical idea presented in the film. For him, this film is "aesthetically revealing" against a world which is replete of "destructive rationalism, reductive instrumentalism and imperialist violence" (Ibid., 182). However, is this transformative for philosophy? Maybe it could be said that this film can do philosophy or that it contributes to philosophy itself, but how should we regard this as *transformative*? It is worth questioning why Sinnerbrink believes film possesses a unique quality that enables it to transform the power dynamics between philosophy and the arts. Why not, for instance, conceptual art? We could argue that conceptual arts, for instance, have already challenged these dynamics and boundaries by provoking reflection on what art is – a typical question of the philosophy of art.

Sinnerbrink argues that film has a specific temporal dimension unique to cinematographic experience – one that endures as we endure, unlike non-temporal art forms. This experience has a specific "charm" which is not captured by "rational reconstructions" nor "cognitivist comprehension". It has a unique aesthetic dimension based on photography, which "exposes the

viewer to an intensive experience of aesthetic singularity" (Ibid., 54). However, we could ask again: could not the same be said about all art forms? All of them could be considered as having an "aesthetic singularity", something specific, even if this is debatable. Moreover, in this case, we can again ask what precisely film has that allows it to transform philosophy. This question will be approached again in the following chapters.

1.3 Revising the dichotomy

Still, for Sinnerbrink, the way to defend the claim that film can do philosophy is through the film-philosophy approach. As discussed earlier, this conclusion is the result of a dichotomy between, on the one hand, philosophy of film which seems to be identified with the cognitivist, analytical, naturalist, rationalist, or Anglo-American theories, and, on the other hand, film-philosophy that defends the idea that film can be philosophy. Similarly, for Martin-Jones, "The most obvious division [is], on the one hand, the exploring of films as philosophy (film-philosophy), versus, on the other hand, more cognitivist and analytical approaches to film (philosophy of film)" (Martin-Jones 2016, 17); and Sorfa states that film "can do philosophy in a way that is unique to the medium" and that "[w]hether this is true of all films, or whether only some films do philosophy, or, whether some films do bad philosophy, is all part of our current research" (Sorfa 2016, 3). So, as Sorfa informs us here, the film-philosophy approach claims that film can be philosophy.

Sinnerbrink seems to justify this approach by presenting a false dichotomy between philosophy of film and film-philosophy in the way they regard film: either you are a philosopher of film, and you disenfranchise its intellectual power, privileging the theorising of film over its aesthetics (Sinnerbrink 2011, 180), or you are a film-philosopher and accept that film philosophises. The first side of the dichotomy seems strange since the theorising of film does not exclude attention to its aesthetics. Noël Carroll has countless works, such as *The Philosophy of Motion Pictures* (2008) or *Theorizing the Moving Image* (1996), where he also seeks precisely to address questions about the aesthetics of film, such as the issue of the cinematographic *medium* specificity, how film can be viewed as an art form or particular issues such as narrative or montage. Carroll provides several tools through which one can make theories about film, such as the paradox of fiction, and through which films can be aesthetically evaluated. Furthermore, besides Cavell, Deleuze, and Sinnerbrink himself, other philosophers identified as film philosophers include Stephen Mulhall, Daniel Frampton, and Thomas Wartenberg. (Cf. Ibid., 120). While it is possible

to understand how some of these could be regarded as film-philosophers, Wartenberg seems to fall into Sinnerbrink's "philosopher of film" category and not the "film-philosopher" category. Wartenberg's arguments and methods to defend film as philosophy are closer to those used by Noël Carroll, whom Sinnerbrink does not include in film-philosophy.

While we might agree with the idea that a division between analytic and continental is not helpful and may even be counterproductive to film as philosophy, the approach that is proposed as an alternative by Sinnerbrink seems to fall in the same kind of dichotomy anyway, even by distinguishing philosophers of film from film-philosophers. From Sinnerbrink's ideas, we would get the impression that only film-philosophers, attentive to film's aesthetics, maintain that film can philosophise, and philosophers of film disenfranchise film and do not accept film as philosophy. Maybe it would be more helpful to use the terms "film-philosophy" and "film as philosophy" interchangeably, not dividing between analytical and continental, not between "philosophers of film" and "film-philosophers", since all philosophers can defend the idea of film as philosophy without having to accept Sinnerbrink, Sorfa or Martin-Jones' versions of film-philosophy.

1.4 Mulhall's case for films as philosophical works

While Sinnerbrink raises important concerns about the disenfranchisement of film in philosophical discourse, Stephen Mulhall approaches the relationship between film and philosophy from a different angle. Rather than focusing on broad theoretical divisions, Mulhall examines specific films to demonstrate how they can engage in philosophical reflection. His analysis of the *Alien* tetralogy provides a concrete example of how films can contribute to philosophical discourse while maintaining awareness of the concerns about disenfranchisement that Sinnerbrink raises.

In the first edition of his work *On Film*, Stephen Mulhall begins his project by analysing the *Alien* tetralogy comprised of *Alien* by Ridley Scott (1979), *Aliens* by James Cameron (1986), *Alien 3* by David Fincher (1992), and *Alien Resurrection* by Jean-Pierre Jeunet (1997).[1]

According to Mulhall, the central philosophical theme in these films is the relationship between human personal identity and embodiment, a central theme to philosophical reflection since Descartes (Mulhall 2001, 2). According to the philosopher, these films should be viewed as

[1] The general narrative of these films focuses on the confrontation of the Lieutenant Ellen Ripley and her crew with an extremely aggressive and dangerous alien species.

genuine contributions to philosophy because of the sophisticated and self-aware manner in which they present themes of philosophical interest. (Ibid., 2). Notoriously, the British philosopher stated:

> I do not look to these films as handy or popular illustrations of views and arguments properly developed by philosophers; I see them instead as themselves reflecting on and evaluating such views and arguments, as thinking seriously and systematically about them in just the ways that philosophers do. Such films are not philosophy's raw material nor a source for its ornamentation; they are philosophical exercises, philosophy in action – film as philosophising (Ibid., 2).

This way, Mulhall presents us with three possibilities about the relationship between film and philosophy:

(1) film as doing philosophy;

(2) philosophy of film;

(3) film in the condition of philosophy. (Cf. Ibid., 130).

The first way of relating philosophy and film (1) is to see film *x* as producing philosophy about specific themes, such as personal identity, as identified by the philosopher.

About the second type of relationship (2), the philosophy of film, Mulhall recalls that this is the most typical or recognisable form of relationship between philosophy and film, being a common practice of philosophy to relate to other human practices that do not usually reflect about themselves or their presuppositions and concepts – philosophy, as he designates, as a parasitic discipline. This would be the form of a relationship between film and philosophy that Sinnerbrink identifies as provoking the disenfranchisement of film.

The third way of relating philosophy and film is film itself in the condition of philosophy (3) (Ibid., 5), a relation Mulhall identifies in the *Alien* tetralogy. According to Mulhall, these films are involved in the philosophy of film because they question their conditions of possibility (Ibid., 2-3). This questioning results from the fact that they belong to a franchise; that is, under the condition of having to make several sequels, the filmmakers had to reflect on the characteristics of the *medium* and the conditions of possibility of the previous films. According to Mulhall, this kind of self-reflection of the medium resembles modernist art: its tendency to critically examine its own conditions. In this sense, the reflective logic of the *Alien* sequels mirrors the way modernist works in literature or painting often questioned their own limits and nature.

Thus, for the philosopher, just as other arts that in modernism achieved the "condition of philosophy" through self-reflection, so does film, especially when involved in a franchise. Now, one must wonder whether belonging to a franchise really makes films question their conditions of possibility in similar ways as modernist paintings – it does not seem to be the case necessarily. Usually, belonging to a franchise just forces new additions to be narratively congruent with previous instalments.

Stephen Mulhall poses the question often posed against these ideas: are not the philosophical ideas "found" in films over-interpretations that the philosopher attributed to those films? In response, the philosopher suggests that these interpretations should be discussed on a case-by-case basis and that several particular interpretations should be confronted to understand whether any of them are closer to what the films try to transmit to us.

Mulhall frequently receives criticism for over-interpreting films, with some claiming that movies cannot address philosophical topics of significance. This "would suggest a rather impoverished conception of the intellectual powers of film and the pervasiveness of matters of philosophical interest in human life" (Ibid., 8-9). The issue at hand is not just the undervaluation of the intellectual capacity of films but also the underestimation of philosophy and its ability to disseminate. Mulhall thinks that individuals in the film industry and others are more interested in philosophy than philosophers might assume. In the opening pages of his book, the philosopher says that he noticed the tendency of cinema theorists:

> To treat the films they discuss as objects to which specific theoretical edifices (originating elsewhere, in such domains as psychoanalysis or political theory) could be applied. Even the most useful of these discussions would usually begin with a long explanation of the relevant theory and turn to the specific film only at the end and only as a cultural product whose specific features illustrate the truth of that theory (Ibid., 7).

Thus, according to Mulhall, film theorists tend to confirm their pre-existing theses in the films, not allowing for greater independence of the film itself. The references made by Mulhall in this excerpt show his disagreement with how cinema is treated by the "Grand Theory", also criticised by Noël Carroll and David Bordwell. For this reason, the philosopher seeks to introduce a new way of theorising about film that prioritises the particular.

Mulhall admits to owing his inspiration to Cavell:

> The possibility that at least some films might be thought of as standing in some internal relation to philosophy is not something I regard as my discovery. I first acquired a sense of its reality and of its applicability to my experience of film through the writings of Stanley Cavell." (Mulhall 2008, 129).

For Stephen Mulhall, it should not be presumed *a priori* that no film is capable of doing philosophy, nor that all films are capable either. Instead, we should look at particular films to see if at least some of them can do it. He also notes that, with this idea that he considers to be modest, he does not mean that most films are capable of this kind of philosophical reflection (Ibid., 131-32). That is, in the philosopher's consideration, whether or not a film does philosophy implies whether her "claims to identify such moments [in which the film is doing philosophy] in these films are convincing" (Ibid., 133). Thus, one of his main criticisms of sceptical philosophers is that "[r]ather than allowing their experience of particular films to teach them what film might be, they permit their preconceptions about the nature of film to dictate what their experience of particular films might be" (Mulhall 2008, 145).

Mulhall's approach thus represents a middle ground in the film as philosophy debate, advocating for a case-by-case evaluation rather than sweeping claims about the film's philosophical capabilities. His emphasis on analysing particular films in order to show their philosophical potential, rather than imposing preconceived limitations, provides a framework that avoids radical positions, whether favourable or sceptical.

1.5 Wartenberg's approach to film as philosophy

Thinking on Screen: Film as Philosophy, published in 2007 by Thomas Wartenberg, greatly influenced the debate about film as philosophy. In this book, Wartenberg clarifies what he considers to be film as philosophy and what it is not. Firstly, it is not the same as "philosophy *in* film".[2] With *philosophy in film*, it could mean that philosophy is present in a particular film, but this way of speaking would not clarify the specific question to be dealt with. *Philosophy in film* might mean that a film presents philosophy: an example of this would be, let us imagine, a recording of a philosopher giving a lecture. Wartenberg's example would be an imagined film called *Justifying Difference,* which would be a recording of a lecture by John Rawls (Wartenberg 2006b, 128).

[2] Expression used by Richard Eldridge (Eldridge 2009, 109).

A real example of a film that does not philosophise is Richard Linklater's *Waking Life* (2001), since the film records philosophy being done verbally (Idem 2016).[3] It is not this that philosophers who defend film as philosophy want to say since, in this case, *Waking Life* does nothing to add to the presented philosophy. In addition, the philosopher also clarifies that the proposal is bolder than saying that films can raise philosophical questions – it is instead that there are films that philosophise on their own.

Tal S. Shamir gives some examples of this, too. For example, in a recorded lecture of a philosopher explaining Descartes' evil genius, "although the audience sees and hears the philosopher, it still has to imagine the entire action of the evil genius narrative, rather than seeing it" (Shamir 2016, 112). In this case, it is not the film itself that would be contributing to the philosophy being done, in this case, the thought experiment. It would be the imagination of the thought experiment guided by the philosopher, not the film itself, that would be doing the philosophical work. Another example he gives is the film *The Cave: An Adaptation of Plato's Allegory in Clay* (Michael Ramsey, 2008).



Figure 1: Still image from *The Cave: An Adaptation of Plato's Allegory in Clay* (directed by Michael Ramsey, 2008)

In this case, the audience can actually have a visual experience of the cave, and it is "not merely confined to the imagination." However, here the philosophical content is expressed mainly by voice-over and "using voice-over to express the philosophical content requires the audience to imagine the philosophical content, rather than experiencing it" (Ibid 113). Thus, in all these cases, it seems that it is not cinema as a whole that is philosophising, and especially not in any original way.

[3] In the film, Robert Solomon appears giving a lecture.

So, what is at stake is whether there is another way to philosophise beyond traditional media (dialogues/orality and written texts), and in this case, the *medium* of film (Wartenberg 2006b, 128), and if so, about what, and how? Wartenberg introduces an important distinction: while certain films can engage in philosophical thinking, he does not equate this with declaring such films to be works of philosophy per se.

While acknowledging that there are films that can be considered works of philosophy, such as the BBC series by John Berger, *Ways of Seeing* (Ibid. 2007, 13), this way of describing the question should not be used for the philosopher. This is due to his idea that for something to be considered a work of philosophy, it is necessary that it is present in the current tradition. Something is a work of philosophy if it has a role in its history and tradition. Thus, a cinematographic work may be considered a work of philosophy, but this cannot be determined immediately. Instead, for Wartenberg, this recognition must come out of the development of philosophy of film itself. However, the philosopher remains sceptical of this possibility since "philosophers are, after all, jealous of their professional identities" and hardly accept the recognition of works of philosophy outside the conventional media (Ibid., 132). Thus, this way of putting the question as to whether films can philosophise without yet being regarded as works of philosophy is advantageous, according to Wartenberg, since the question of whether works of philosophy have intrinsic characteristics remains an open one, maintaining the discussion about how certain films can philosophise (Ibid., 131).

In addition to these considerations, Wartenberg also explains why he considers film a possibility for doing philosophy. For him, "taking movies seriously is one way to help people see the importance of philosophy" (Ibid., 8). Here, he presents a concern for the very role of philosophy in contemporary society, which, in his view, should be more relevant. For him, a society that wants to be more rational can benefit from greater accessibility of people to philosophy, and this can be done through popular culture, especially through Hollywood cinema. This connection to film can even be a means to ensure the "vitality" of philosophy and recognise its importance in today's societies (Ibid., 141-42).

Despite these motivations, Wartenberg starts his work with a sceptical, although enthusiastic, approach. Thus, with caution, the philosopher determined that his strategy should be to adopt a modest position and examine which techniques or tools of philosophy are possibly also shared by film.

One tool or technique that Wartenberg suggests film can use in order to philosophise is the illustration. One of the main problems Wartenberg points out about *On Film* is how Mulhall unreflectively rejects illustration as having any relevance to philosophy. Wartenberg questions whether using the expression "mere illustration" does not carry preconceived and *a priori* notions about the value, or lack thereof, of illustrations. Despite his rejection of presuppositions about philosophy, proposing that film can also show us what philosophy can be, Mulhall makes the same mistake he identifies, according to Wartenberg. The dichotomy that Mulhall creates between "handy illustrations" and "serious and systematic philosophical thinking" concerns Wartenberg about *On Film*. In addition to the problem of not questioning illustration as just something "handy", the other part of the dichotomy is also problematic.

Suppose Mulhall identifies philosophical thought with systematicity and seriousness. In that case, there are several counterexamples to this: Wartenberg provides those of Kierkegaard, Nietzsche, and Wittgenstein against the idea of systematicity, and on the other hand, he refers to Derrida and his idea that not all philosophical thinking has to be "serious" (Cf. Ibid. 2006a).

Another philosopher who defends film as philosophy and makes the same dichotomy between mere illustrations and valuable philosophy pointed out by Wartenberg is Christopher Falzon in his work *Philosophy Goes to the Movies: An Introduction to Philosophy* (2002). In this book, the philosopher distinguishes "the use of films that 'illustrate' philosophical positions" from exploring philosophical content in the films themselves (Ibid. 2002, 20).

Wartenberg points out several issues with this distinction made by Falzon. The first is that the philosopher does not explain why his analysis of certain films treats them as more than illustrations (that is, he does not make an independent argument to show that these films are more than illustrations). Second, Wartenberg suggests that Falzon's treatment effectively reduces films to visual representations of pre-existing philosophical concepts, thereby denying them any genuine independent philosophical potential. (Ibid., 21)

Thus, the way Falzon tries to justify the idea that the films he analyses philosophise is that philosophy and film use images. In this way, Falzon could try to explain how the use of images in films can be a way by which they philosophise. What the philosopher affirms is, however, that his "interest in images is not so much in the role of the image in philosophy as in the philosophy we can discern in the image. So, my concern here is indeed with the image insofar as it is illustrative

and captures something about philosophical thought at a concrete level. However, even here, the image is not merely illustrative." (Ibid., 5).

So, here, Wartenberg identifies a contradiction in Falzon's ideas: on the one hand, he states that certain films merely illustrate philosophical ideas; on the other hand, he states that philosophical texts use images to illustrate their philosophical positions. Thus, if philosophical texts (or works of philosophy) use illustrations for philosophical theses, why deny this to film?

Wartenberg seeks to provide some reflections on what illustration itself is and what kinds of illustration may exist.

1.6 Film as philosophical illustration

The first idea is that illustrations have intentionality; they always refer to something beyond themselves. He categorises the illustrations into four types:

1) *Illustration that is subordinate to the text*. In this case, characters and events, the narrative, and the imaginary depend on the text, and illustration supports the narrative. Examples of this are illustrations of Mark Twain's *The Adventures of Tom Sawyer*, which only supplement the text to make it more accessible – the text itself is not "substantially enriched by illustration" (Wartenberg 2006b, 21).

2) *Iconic representations* or illustrations that work with the text in our imaginations of these worlds. One example is *Alice's Adventures in Wonderland*, *Harry Potter* or *Winnie-the-Pooh* – "The fictional world built by the book depends on both the text and its illustrations" (Ibid., 25).

3) *Illustration that is essential and indispensable to recognise what is described in the text*. For example, animal illustration books such as the *Handbook of Birds of the Western United States* in which there are paintings of the birds and a description thereof, which serves to identify the illustrated animals. (Ibid., 26-27).

4) *Illustration that becomes an independent work of art*. For example, the illustrations made by Marc Chagall for the story of *Daphne and Chloe*) (Idem 2006b, 43).

Wartenberg shows that not all illustrations are subordinate to what they illustrate.

After these considerations, Wartenberg seeks to question whether there are examples of films that philosophise in this way by illustrating a philosophical theory.

Wartenberg analyses John Ford's *The Man Who Shot Liberty Valence* (1962) as an illustration of Nietzsche's criticism of Hegel's view of History (Ibid., 54). The way the philosopher

argues for the idea that this film illustrates a Nietzschean critique of Hegel's thesis on the progress of History will not be elaborated on in this dissertation, significantly since his idea of the film has changed. For Wartenberg, the film deals not with the progress of history but instead with a contrast between the old West, which lives based on communities, and the modern West, which is organised in societies. The contrast is between community life as an organisation with affectionate and interdependent personal relationships and life in a society where relationships are more distant and based on benefits. The idea, then, is that "Law and Order" interrupted relations that were real. That is, the film may instead be seen as a critique of modern society that has not progressed (not in interpersonal aspects).[4]

Wartenberg's most crucial example of a film that illustrates philosophy is Charles Chaplin's *Modern Times* (1936), a "humorous but philosophically significant" illustration of Marx's theory of alienation (Idem 2006a, 27).

This film begins by juxtaposing two images in a "symbolic montage": a group of sheep walking toward the slaughterhouse and a group of workers walking toward a factory.

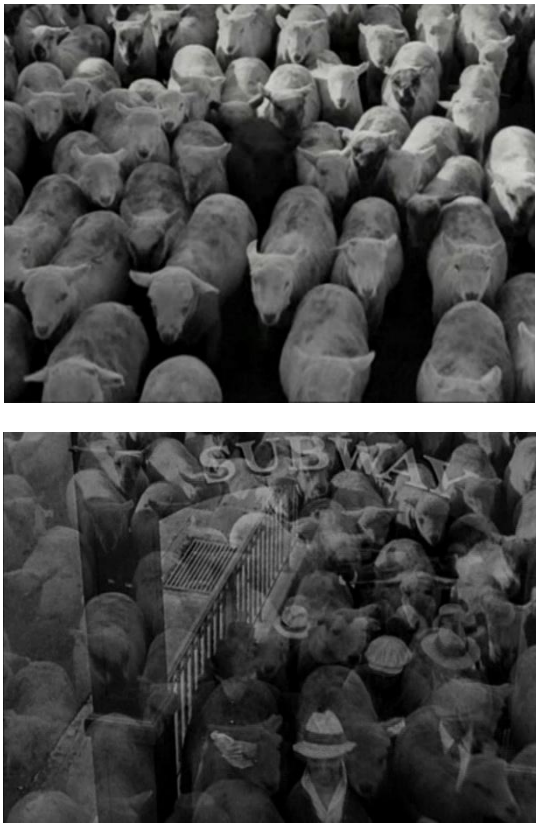


Figure 2: Images from Annette Bochenek, "Smile: A Reflection on *Modern Times*," *Hometowns to Hollywood*, accessed April 26, 2025.

[4] Presentation of the communication "The Death of Love: *The Man Who Shot Liberty Valance*'s Dramatization of Philosophy" at the International Conference on Moving Image and Philosophy at the Faculty of Arts and Humanities of the University of Porto, 2017.

For Wartenberg, this image (or juxtaposition of images) is a concretisation of Marx's idea that workers are dehumanised in a capitalist society and become tools of the economic system itself (Ibid., 28).

Chaplin manages to materialise this idea in several ways, in this juxtaposition of images, but especially in scenes where his character works on the assembly line. In these scenes, Charlie's job is to fasten screws on the assembly line at a relatively fast and very repetitive pace. What happens, humorously, is that even when not on the assembly line, Chaplin cannot stop doing the same movements as he sees anything that is more or less similar to the screws. This is because Chaplin became a "machine" by doing that job.

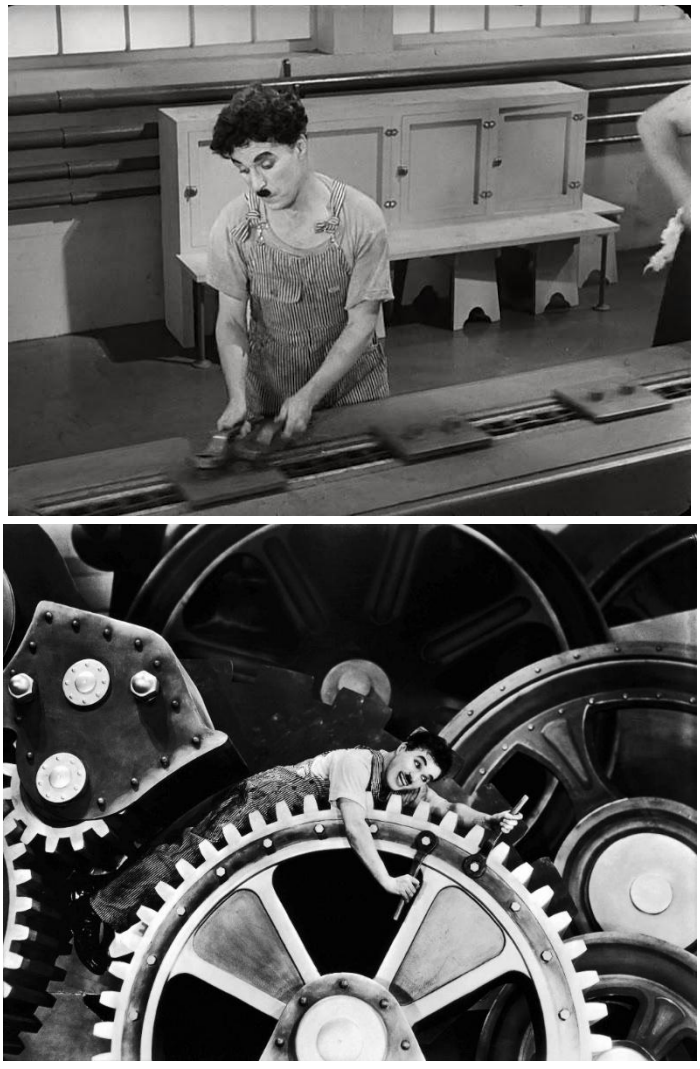


Figure 3: Images from Annette Bochenek, "Smile: A Reflection on *Modern Times*," *Hometowns to Hollywood*, accessed April 26, 2025.

Thus, according to Wartenberg, this film specifically or concretely presents Marx's abstract idea of human mechanisation. Here, human beings have become an "appendage of the machine," and their needs are irrelevant – the capitalist's sole objective is to profit and, to that end, only the rate of rapid production matters, independent from the workers' interests. In short,

In this context, comedy may be especially effective in communicating this critique, since it allows us to confront painful and absurd realities with some distance that comedy films allow. Comedy often reveals the arbitrariness of social structures, and by "laughing at power, we expose its contingency". Thus "'True' humour changes the situation, tells us something about who we are and the sort of place we live in, and perhaps indicates to us how it might be changed" (Critchley 2002, 11). In *Modern Times*, Chaplin's mechanical movement and breakdowns on the assembly line embody this kind of subversive humour, highlighting the dehumanising effects of industrial capitalism through exaggerated, absurd action. Instead of presenting Marx's critique "seriously", *Modern Times* encourages viewers to reflect through humour, making the critique more approachable and less abrasive.

This dehumanisation and mechanisation of humans are also illustrated in other ways. An example that Wartenberg refers to is the "Billows Feeding Machine," the machine that Charlie's boss tries to implement to feed the workers more quickly and, thus, give more time for the workers to produce. This scene exemplifies what Marx called the "commodification of the worker," the idea that workers are treated only as potential sources of profit (Ibid., 49).

The film also presents other scenes that seem to illustrate Marx's ideas. An example is the very relationship that Chaplin establishes with a girl who lives on the street. Several times during the film, these characters dream of a "normal life" in which he goes to work and returns to the house where she is waiting for him. For Wartenberg, these scenes show Marx's idea that it is a mistake to think that poor people are lazy; Instead, these people work in such a way that dehumanises them and turn their efforts futile due to the capitalist economic system itself that focuses on profit only (Idem 2006b, 52-53). In addition to these scenes, in his relationship with the girl, Charlie demonstrates his creative side (in a scene where he shows her how to skate) instead of his work life in the factory, which is wholly confined and exempted from any creativity (Ibid., 52).

Thus, Wartenberg proposes that this film, as an illustration of Marx's philosophical theory, must be seen as philosophising. Thus, it may be said also that being an illustration does not prevent it from being philosophy. Hence, according to the philosopher, "such cinematic illustrations of

philosophical theories play an important role in transmitting the ideas developed by philosophical theories to a wider audience" (Ibid., 53). Finally, according to Wartenberg, one should not assume that spectators only enjoy this film for its humour and not for its philosophy. Instead, the film is comical precisely because the images of Chaplin's mechanisation are linked to its Marxian idea – the comedic value comes from this very idea. In other words, the film's philosophical content is not separate from but integrated with its comedic elements – the humour arises from the visualisation of abstract philosophical concepts about alienation and mechanisation, making the philosophical critique more accessible than a purely theoretical exposition.

1.7 Structural film and self-reflection

While Wartenberg demonstrates how films can philosophise through illustration, as seen in *Modern Times*' engagement with Marxist theory, Noël Carroll identifies another distinct way that films can contribute to philosophical discourse: self-reflection. This approach is particularly evident in structural films, which explicitly examine the nature of cinema, moving beyond illustration to engage in direct philosophical investigation of their medium.

Structural film appeared between the sixties and eighties, framing itself as experimental and avant-garde. As defined by P. Adams Sitney, a prominent avant-garde film historian, this cinematic approach prioritised the film's structural elements, compelling viewers to concentrate on the work's formal composition rather than its narrative or representational content (Sitney cited in Wartenberg 2006b, 118).

What the filmmakers of these experimental films sought to do through their cinema was to expose the fundamental elements of a film. To expose what essential cinema is, these filmmakers, including Michael Snow, Hollis Frampton, Ernie Gehr, Tony Conrad, Paul Sharits, and Andy Warhol,[5] sought to abstain their films from any content considered non-essential to the cinema, such as a narrative, provoking spectators to reflect on the nature of the film: "filmic reflexiveness is forced through cinema's materialist operations of filmic practice" (Gidal 1978, 10).[6] Any content, such as the narrative or characters, was taken away from films, or, in the case of its existence, it was "subordinated to the films' greater concern with its overall shape and structure" (Wartenberg 2006b, 118).

[5] For Sitney, Warhol was the pioneer of the structural film (Cf. Sitney 2002, 349).
[6] Peter Gidal, theorist and historian of structural film, gives also this kind of cinema the designation "materialist film" since this kind of cinema tries to reduce the film to its materiality, or, repeating the same idea, to the fundamental material elements of film (Cf. Gidal 1978).

To expose these fundamental elements of cinema, the filmmakers did the opposite of what mainstream cinema, such as Hollywood cinema, or in general, narrative cinema does. Mainstream cinema does not expose its techniques; quite the opposite, they hide how cinema does what it does – this is called the "transparency" of narrative cinema (Cf. Ibid.). Finally, structural film had only its conditions of possibility and techniques as its subject matter. We could think of what structural films do as instances of what Nelson Goodman calls "exemplification".

According to Goodman's aesthetic theory, a symbol is a tangible object that exemplifies particular properties or characteristics (Goodman 1976). Goodman posits that symbols can be comprehended based on their exemplification or expression of specific characteristics. If a symbol exemplifies particular properties, those properties become significant to it as a symbol. In this context, exemplification refers to the connection between a symbol (like a piece of art) and the properties it exemplifies or expresses. To put it differently, when a work of art exemplifies certain traits, those traits become pertinent to it as a work of art.

Filmmakers in structural film place emphasis on the physical attributes of the film medium. They highlight the characteristics of celluloid film, such as the flickering of frames, the texture of the film stock, and the mechanics of projection. Drawing from Goodman's ideas on exemplification, these formal elements serve as examples or representations of the essential qualities of the film medium. The film's materiality becomes a symbol for the medium, and through exemplification, it conveys the fundamental nature of cinema as that medium.

To understand how structural films could be seen as exemplifying their medium, it is crucial to understand Goodman's distinction between denotation and exemplification. While denotation is a one-way relationship where a symbol refers to something (like a painting of a landscape denoting that landscape), exemplification works bidirectionally – the symbol both possesses and refers to its properties. Tony Conrad's *The Flicker* (1966) could illustrate this concept. The film consists entirely of alternating black and white frames at varying frequencies, creating a stroboscopic effect that exemplifies the fundamental mechanism of cinema: the rapid succession of still frames creating the illusion of movement and light. Unlike conventional self-reflexive cinema, where filmmaking techniques might be referenced or shown but are not necessarily exemplified, *The Flicker* does not merely represent the properties of film projection – it embodies them. The flickering light is not depicted or represented; it is physically present, turningthe film into what Goodman would call a 'sample' of its properties. Conrad's work demonstrates how exemplification in structural film operates by making the medium itself the message. The film's form becomes

inseparable from its content, as it simultaneously possesses and draws attention to cinema's basic material conditions.

According to these ideas about structural film, Noël Carroll claims to prefer to designate this type of cinema as minimalist since these films are aligned with the artistic movement of minimalism in general on its quest for reflexivity through reduction. A minimalist artwork may intend to disclose essential features of the art form or medium it inhabits, and/or it can be designed to promote apperceptive reflection on the part of the viewer concerning the phenomenological dimensions of their encounter with the work, including the way the work modulates, or conditions said experience. The artwork acts as a catalyst for this self-awareness by revealing important aspects of the art form or by prompting the viewer to reflect on their own experience of the artwork and how it affects their perception.

Moreover, the minimalist work provokes reflection on itself and/or its associated experiences by employing highly reductive or limited means (Carroll 2006, 176).

According to Carroll, these types of films are, thus, "Meta-Cinema"; they question and seek to answer what essentially cinema is. For Carroll, Ernie Gehr's *Serene Velocity* (1970), in particular, is possibly the central or most representative work of structural film that philosophises. Through its exploration of the essential elements of a film, it produced a philosophical position on what cinema is.

This film consists of a single, continuous shot of a corridor in an empty institutional building, with the camera slowly zooming in and out at alternating speeds. The contrast between stillness and movement prompts the viewer to contemplate the distinction between photography and filmmaking. Unlike photography, which captures a single moment in time, filmmaking can create motion as its fundamental element. Through his experimental film, Ernie Gehr sought to make an "ontological claim about the nature of the moving picture" (Ibid.), providing his film as evidence.



Figure 4: Image from "Cinemateca Portuguesa: Os Filmes de Ernie Gehr," Experimental Cinema, accessed April 26, 2025.

Thus, according to Noël Carroll, this film

> It is an instance of philosophising through moving images. Its topic, the nature of the motion picture, is unquestionably philosophical. This is undoubtedly not the only question that a philosopher may ask of motion pictures, but it is undeniably one of them. Moreover, *Serene Velocity* proposes its answer to that question using the art of the motion picture — by juxtaposing settings on the zoom lens (Idem., 2006, 179).

According to the philosopher, viewers informed about these themes understand what the film tries to tell us. There is no need for an explanation "in the same way that professional philosophers usually do not need to be told how a thought experiment does its work" (Idem., 2017, 278). Thus, for Carroll, this film functions as a philosophical thought experiment through cinematic rather than usual textual means (Idem., 2006, 178).

Given that until the 1970s, there was no conceptual, systematic, and argued analysis of the philosophical nature of cinema as moving image,[7] this was a very original and important contribution to the philosophy of film. In addition, Carroll states that this film, and several films of the same genre,[8] may be seen as substantial contributions to philosophy. In the philosophy of film, there is a plethora of philosophical ideas derived from those kinds of movies that can aid in understanding the essence of cinema. Thomas Wartenberg shares Noël Carroll's belief that structural films can be a valuable resource for philosophical films. He acknowledges that while essential elements in narrative films often go unnoticed, structuralist filmmakers purposefully expose them to the viewer. This allows us to focus on the structure and purpose of the film rather than just the narrative. While Wartenberg acknowledges that structural film should be viewed within the context of the artistic movements of its time, he does not believe it is essential to classify Serene Velocity as a minimalist film, which Carroll suggests. Minimalism sought to eliminate all elements

[7] Only in 1979, with the article "Moving Pictures" of Danto.
[8] Some other films that can be included in this genre are, for example, *Wavelength* (1967) by Michael Snow, *The Flicker* (1966) by Tony Conrad, *T,O,U,C,H,I,N,G* (1968) by Paul Sharits, or *Zorns Lemma (*1970) by Hollis Frampton.

"superfluous" to the essence of the art form; structural cinema did the same. Still, for Wartenberg, this should not be a reason to change the terminology: "The term *structural* has deep validity in characterising the nature of these films. Given its nearly universal acceptance, there is not sufficient justification for jettisoning it" (Wartenberg 2006b, 120).

Despite agreeing with Noël Carroll that *Serene Velocity* and other structural films may be cases of philosophy through film, Wartenberg disagrees with the way the philosopher states how these films contribute to philosophy: according to Wartenberg, structural films are not, as Carroll claims, thought experiments, mainly because they do not serve to test any pre-existing philosophical theory as it happens in philosophical and cinematic thought experiments. These films are, instead, actual cinematic experiments that may even fail to illuminate cinema's essence if they subtract too many elements from the film, such as narrative, sound, movement, etc., potentially crossing a threshold where they no longer function as meaningful instances of the medium they seek to define.

They attempt to demonstrate which traits are essential to the cinematic medium's nature. They help philosophers understand this nature, and thus they are philosophical films (Ibid., 2006b, 121-21).

Thus, to look further into how structural films can philosophise, Wartenberg goes back to Arthur Danto's ideas about Andy Warhol's cinema in "The Philosopher as Andy Warhol." For Danto, "Warhol's films 'make an unparalleled contribution to our understanding of the concept of film' (Danto 1999, 65). In *Empire* (1965), Warhol subverts the traditional notion that films primarily document object movement. Danto interprets this work as a philosophical critique that exposes the limited conception of cinema as a medium of physical motion. This idea that it was essential for cinema to present moving objects derived from Rudolf Arnheim (1957). Arnheim's theoretical position posits movement as the essential characteristic distinguishing cinema from other visual arts. His argument rests on the assertion that film's fundamental artistic identity and power derive from its unique capacity to present motion, transforming static visual elements into dynamic temporal events. His theoretical framework emphasises how the technical process of creating motion through sequential frame projection enables unprecedented manipulation of temporal and spatial relationships. This conception of cinema's artistic essence as fundamentally rooted in movement would later be directly challenged by experimental works like Warhol's *Empire*, which deliberately subverts these assumptions about cinematic motion through its sustained presentation of a stationary subject. During the eight hours of the film, the presented object, the Empire State Building, does not move.



Figure 5: Image from "EXIBIÇÃO EXCLUSIVA: 'Império' de Andy Warhol completa 60 anos!", Empire State Building, accessed April 26, 2025.

For Wartenberg, Empire can thus be seen as a counterexample to this definition of cinema as capturing moving objects. In this way, Wartenberg concludes why this film can be considered philosophically rich:

> Warhol has here discovered a mode of filmmaking that transforms the viewer's experience from one in which she is focused upon the pictorial and narrative content of a film into one in which she also notices those aspects of the film that are normally transparent, but that are necessary for a film to present any object whatsoever to its viewers (Ibid., 127).

Empire is so significant because "Warhol has discovered an analogue to Kant's transcendental philosophy" (Ibid.) by questioning what structures are necessary for cinema to make an object appear on the screen, and this is why Wartenberg designated this film as a "work of *transcendental cinema*" (Ibid).

Wartenberg returns to Tony Conrad's *The Flicker*, previously mentioned, to show how it philosophises in a way that goes beyond *Empire*. While *Empire* was still, for Wartenberg, based on an assumption of photography as essential to cinema – an idea advocated by André Bazin in *The Ontology of the Photographic Image* (1960) – Conrad's work represents a more radical criticism of that assumption.

Conrad's film *The Flicker* shows that the camera does not have to pick up moving objects to originate moving images photographically – the film technique itself can produce a motion picture without filming moving objects. "According to *The Flicker*, *this* is necessary for something being a film: its employment of the technology for stroboscopic image projection" (Ibid., 130). In addition,

> *The Flicker* shows its audience its cards: the flicker. It is as if it were saying, "See this? Well, I'm going to produce motion from the simple alternation of an illuminated and a darkened screen. There is no more to cinema than that!" (Ibid., 131).

These films, for Wartenberg, can thus be considered works of art, in this case, cinematographic works and potentially works of philosophy simultaneously (Ibid., 132).

A defence of the idea that structural films can philosophise can also be found in Jinneeh Choi (2006). This film theorist points out that structural cinema, and avant-garde in general, is valued for its originality (because it advances new hypotheses and possibilities for the medium) and consistency (the hypothesis posed by the film must be consistent with what the movie shows). For Choi, this cognitive value that structural films have adds to the overall value of the film, including its aesthetic value – so for the theoretician, if someone claims to have understood the film but did not appreciate it, it means that either the film was not coherent or consistent – that is, it was not able to show or prove the statement it intended – or it is the viewer herself who did not understand the film.

This chapter has traced the development of theoretical approaches to the relationship between film and philosophy, examining how various theorists have articulated different ways films can be philosophical. The exploration began with Robert Sinnerbrink's critique of existing approaches, particularly challenging the dichotomy between "Continental" and "Analytical" philosophies of film. Sinnerbrink defends the "film-philosophy" approach that treats film as an autonomous philosophical medium, capable of philosophising by itself. This analysis incorporated perspectives from Cavell and Deleuze, who viewed film as having a unique relationship with philosophy. Cavell saw cinema as a means of addressing philosophical scepticism by retrieving and validating ordinary experience, while Deleuze understood film as a potential generator of entirely new ways of thinking.

Then, an investigation of multiple perspectives on how film might philosophise was made. Stephen Mulhall tries to show that films may philosophise in serious ways "just like" traditional philosophy, through his analysis of the *Alien* tetralogy. Thomas Wartenberg proposed illustration as a legitimate philosophical method, using Charlie Chaplin's *Modern Times* as an example of how a film can illustrate philosophical concepts – in this case, Marx's theory of alienation. Noël Carroll and Wartenberg's exploration of structural film offered another approach, highlighting how film can philosophise by examining the nature of its own medium. Films like *Serene Velocity* and *Empire* were presented as instances of film that philosophically interrogate the fundamental conditions of film.

Throughout these discussions, there was an approach to the idea that film is not just a subject of philosophical analysis, but it may be a medium capable of philosophising by itself, in some way. In the following chapters, I will argue how specifically films may be better or not at philosophising.

2. Preliminary objections against film as philosophy and the appeal to thought experiments

As established previously, advocates of "film as philosophy" (FAP) contend that films are capable of philosophising in various ways. This perspective, however, has been met with some resistance from critics. Objections against FAP often focus on the nature of film as an art form, its reliance on the narrative and visual means, and its perceived inability to produce explicit arguments or universal truths.

This chapter will explore these objections and their counterarguments in detail. However, it is worth noting that much of this debate appears to rest on an implicit hierarchy that positions philosophy as superior to dramatised fiction. A question here emerges: what would happen if we questioned this hierarchy? Rather than asking whether film can rise to philosophy's level, we might consider philosophy as waiting for dramatisation somehow. This would suggest that film might not be an inadequate form of philosophy, but rather a medium that can actualise philosophical concepts in ways that traditional means like texts cannot. By analysing these debates, this chapter aims to clarify both the potential and limitations of film as a medium for philosophising.

Some of the main objections to the film as philosophy thesis include the generality objection, the explicitness objection, the banality/triviality objection, and the propriety objection. In response, defenders of FAP have offered several counterarguments. This chapter will explore these objections and their counterarguments in detail.

2.1 The banality objection

Some objections against film as philosophy are closely related to objections against literary cognitivism – the general idea that literature can have cognitive value.

Jerome Stolnitz argues that literature has little cognitive value (Stolnitz 1992). He makes the "banality objection", asserting that literature and the arts in general offer little to no cognitive value, since they do not reveal meaningful or substantive truths about the word.

According to Stolnitz, one of the main issues with the idea of literary truths is that they cannot be contradicted or confirmed, making them epistemically unreliable. Unlike the rigorous methods used in philosophy or science, art is ambiguous and unverifiable. Furthermore, Stolnitz

highlights the absence of experts dedicated to studying artistic truths alone, further highlighting that such truths lack any intellectual expertise and scrutiny (Ibid., 198).

An important aspect of this banality argument is that any significant truths, such as moral lessons about human nature, existential insights about meaning or mortality, or ethical principles about justice or virtue, supposedly conveyed by art in general, and literature in particular, are ultimately banal or trivial. So, the knowledge that apparently derives from works of literature is just a set of truisms. Stolnitz argues that literary works teach nothing novel or profound; instead, they merely reiterate preexisting notions that audiences are already familiar with (Ibid., 200). Even when a literary work presents a significant truth, Stolnitz maintains that this truth is not unique to the art form itself but belongs to "some extra-artistic sphere of the great world" (Ibid., 198). This implies that the insights presented in art are merely reflections on broader philosophical or cultural discourses rather than original contributions to knowledge.

Stolnitz further argues that artistic truths, particularly those found in literature and cinema, are primarily concerned with narrative, character development, and aesthetic expression rather than philosophical insights. As a result, the truths conveyed in artistic works are often vague. He claims that the truths presented in works of art are "preponderantly banal" (Ibid., 200). To illustrate this critique, consider how Stolnitz might approach *The Matrix*. While defenders of FAP argue that the film offers insights into scepticism and the nature of reality, Stolnitz would likely dismiss these as merely repackaging Cartesian doubt. The film's explorations of a simulation would not teach audiences anything that Descartes' evil demon has not already established more rigorously. From this perspective, *The Matrix*'s apparent philosophical depth would amount to little more than spectacular packaging of familiar epistemological problems, confirming that audiences are simply encountering truisms.

This banality critique is linked to Stolnitz's broader theory of aesthetic attitude, which holds that art should be appreciated disinterestedly, without any cognitive or practical aims. From this perspective, the cognitive value of art is not only trivial, but also unnecessary to its purpose (Ibid., 198). If one accepts this view, seeking cognitive value in art may even be considered a mistake, as it imposes an inappropriate expectation onto a medium primarily designed for aesthetic contemplation.

Another dimension of the banality objection is the claim that the audience often presupposes general truths found in art. Noël Carroll elaborates on this point, arguing that films, for example, frequently reinforce ideas that viewers already accept: Thought experiments allow us

to observe the results of the imaginative employment of our antecedently possessed verbal and conceptual know-how in such a way that we become aware of what we already know" (Carroll 2002, 8).

From this point of view, art does not introduce new insights or expand knowledge but merely reiterates commonly held beliefs, reinforcing preexisting ideas rather than challenging or reshaping them. Thus, art is seen as a medium that merely reinforces "truisms that readers already know" (Stolnitz 1992, 198).

Catherine Elgin challenges Stolnitz's objections' assumptions by reconceptualising the relationship between art, science, and understanding. She argues that "because of its narrow focus on the conditions for knowledge, contemporary epistemology cannot say what makes insights interesting or important. Thus, it cannot say what knowledge is worth having and seeking" (Elgin 1993, 13).

Thomas Wartenberg argues that it must be shown how a film can do something non-trivial, as he thinks may be the case with *The Third Man*. Noël Carroll and Thomas Wartenberg offer differing interpretations of *The Third Man* (1949), directed by Carol Reed and written by Graham Greene, as a film with philosophical significance. Carroll views the film as a counterexample to the maxim: "[w]hen loyalty to a friend conflicts with loyalty to a cause, one ought to choose in favour of the friend" (Carroll 2002, 10). He argues that the film challenges this principle by presenting a narrative in which loyalty to a friend leads to morally questionable outcomes. However, Wartenberg disagrees with Carroll's interpretation, asserting that the film is not merely a counterexample to a thesis but a thought experiment that explores the complexities of moral intelligence (*phronesis*) and illustrates Aristotle's approach to friendship. While both philosophers agree that friendship is a central theme of the film, Wartenberg argues that *The Third Man* makes a more substantial philosophical contribution than Carroll considers.

Wartenberg begins by explaining Aristotle's concept of moral intelligence in *Nicomachean Ethics*, which involves the ability to apply ethical values to concrete situations. He argues that *The Third Man* engages with this idea by demonstrating how various factors, such as personal commitments and the romanticised view of life shaped by mass media, complicate moral decision-making. Wartenberg writes, "[o]ne philosophically significant task that a film – as well as works of narrative fiction – can do is to allow us to see the complexity of our moral lives in a manner that will enable us to grasp morally significant aspects of it fully." (Wartenberg 2007, 99). The film, set in post-war Vienna, depicts a city rife with black markets, including one involving the sale of diluted

penicillin, which leads to numerous deaths. The narrative centres on Hollis Martins, a writer of Westerns, who travels to Austria at the invitation of his friend Harry Lime. Upon arriving, Martins is confronted with the apparent death of Lime and, later, the revelation of Lime's criminal activities.

Initially, Martins refuses to believe that his friend could be responsible for such crimes, reflecting Aristotle's assertion that "one of the advantages of friendship is that friends do not believe lies told about their friends" (Ibid., 102). However, this loyalty to Lime becomes a moral obstacle, as Martins' romanticised view of their friendship prevents him from cooperating with the police. Wartenberg identifies this as one of the "vicissitudes of moral intelligence," illustrating how personal commitments and distorted perceptions can hinder ethical decision-making. Martins' eventual realisation that Lime is guilty forces him to reevaluate their friendship. In the end, Martins assists the police in capturing Lime and, in the process, kills him. This act underscores Aristotle's thesis that friendship must be deserved and that loyalty should only be extended to true friends. Wartenberg argues that the film's exploration of these themes contributes to Aristotle's theory of moral intelligence by presenting a complex portrayal of ethical dilemmas.

Wartenberg further contends that *The Third Man* is a thought experiment on the variations that make moral intelligence problematic and complex. By illustrating Aristotle's claim that there are cases where the dissolution of a friendship is morally justified, the film engages with philosophical ideas in an original and effective way. Wartenberg concludes that *The Third Man* "does philosophy in a particularly effective and original manner" (Ibid., 116). He argues that the film's contribution to the theory of moral intelligence demonstrates the potential of cinema to make cognitively relevant and interesting philosophical contributions.

In summary, Wartenberg uses *The Third Man* as an example of how a film can address non-trivial philosophical questions. He argues that the film's exploration of moral intelligence and its illustration of Aristotelian ethics qualify it as a work that is "a real contribution that the film makes to the theory of moral intelligence and thus qualifies this film as doing philosophy in a particularly effective and original manner" (Ibid., 116). This challenges the notion that films are limited to conveying banal or trivial truths, demonstrating instead their capacity to advance our understanding of complex philosophical ideas.

Noël Carroll also responds to this objection by stating that there are, in philosophy, ways of producing knowledge that seek to mobilise knowledge that the audience already has, such as thought experiments, examples and counterexamples: these serve to provoke "reflection on what is already known, especially in terms of the application of concepts; nevertheless, we still regard

them as productive knowledge, since they make what in some sense is already known accessible and salient" (Carroll 2002, 8).

Furthermore, Carroll questions: "banal for whom?" For example, Peter Kivy argues that much literature has as its theme the hypothesis of determinism or fatalism, which can be seen as "philosophical banalities" since such themes are very well-known to academic philosophers. However, Kivy points out that questions about free will are of "deep concern for people" though the philosophical assumptions are not familiar to all readers, given that these works aim to engage a wider audience (Kivy 1997, 20). According to Kivy's observation, Carroll recalls, then, that art is usually directed to a broad audience that is not restricted to an academic philosophical one and, therefore,

> Conclusions that might appear utterly banal or obvious for experts in ethics may not be banal or obvious to nonprofessional audiences (...). What the philosopher discounts as trivial may be revelatory for the plain reader and, for that very reason (...) [the work of art may] be informative and educative for the intended audience (Carroll 2002, 10).

While Carroll's argument here might initially position film as merely popularising philosophy for non-specialists, this understates its significance. The philosophical value of film should not be measured by its ability to generate novel ideas that surprise academic philosophers.

So even if the work were directed at an academic audience, philosophy uses techniques such as thought experiments that resemble literary or cinematic narratives and presuppose that the viewer has the theoretical framework to understand the point of those issues.

Carroll also asserts that mainstream films are usually made for popular consumption. Therefore, certain ideas may become banal since cinema itself has been propagating such ideas throughout time (Carroll 2008, 24).

2.2 The explicitness objection

One of the most persistent objections to film as philosophy is the "explicitness objection." This objection is closely related to what Noël Carroll calls the "no-argument argument" against literary cognitivism, building on observations by Peter Kivy about how literary works express propositions (Kivy 1997, 18). The core claim is that even if works of art suggest universal truths, they neither present explicit arguments nor engage in formal critical debate to support them (Carroll 2002, 6).

Carroll identifies two key claims within the no-argument argument:

1) Artworks assert or imply truths rather than arguing for them.
2) The critical discourse surrounding artworks does not typically focus on debating the validity of these truths.

To clarify the discussion surrounding the explicitness objection, it is helpful to consider arguments made by Lamarque and Olsen. They argue that literary criticism is "not primarily concerned with adjudicating the truth-value of general propositions, but rather with exploring how these propositions are interwoven with narrative structure and aesthetic form" (Lamarque and Olsen 1994, 332). In this view, critics focus on how ideas shape a work's overall organisation and meaning rather than engaging in debates over their literal truth.

Lamarque and Olsen further contend that literature does not function as a repository of propositional knowledge like science or philosophy. They assert that "the primary merit of a literary text lies in its ability to evoke emotional resonance and imaginative engagement, qualities that transcend the need for explicit argumentative precision" (Ibid., 335). This means that these literary qualities operate in a different epistemic domain, where emotional and imaginative responses can generate understanding without requiring the logical rigour demanded by philosophical argumentation. This observation underscores their claim that the evaluative criteria for literature differ fundamentally from those employed in traditional philosophy.

According to Lamarque and Olsen, literary texts are not primarily vehicles for conveying propositional truth. Instead, they generate meaning through suggestion, narrative, and symbolic interplay. They state, "[l]iterary works function by implying truths rather than stating them overtly; the absence of a formal, logical argument is not a shortcoming but an intrinsic characteristic of the medium" (Ibid., 337). This feature could be reframed as a significant advantage: literature's indirect approach may enable it to address philosophical questions that resist direct argumentation, which is an argument to be developed for the case of film.

Moreover, they emphasise that the evaluative framework for literary art should recognise the inherent differences between art and philosophy. While philosophers are typically assessed based on the clarity and rigour of their arguments, literary critics are more concerned with how effectively a work's narrative, imagery, and stylistic choices contribute to an overarching theme. Lamarque and Olsen note, "A novel or film may allude to existential dilemmas and moral ambiguities without ever systematically arguing for a single, definitive thesis. The value lies in its capacity to mirror the complexities of human experience through its form" (Ibid., 340).

In film as philosophy, the explicitness objection states that film "lacks the explicitness to formulate and defend the precise claims characteristic of philosophical writing" (Wartenberg 2007, 16). Wartenberg traces this objection to Plato's critique of art in *The Republic*. In Book X, Plato describes art as mere imitation (*mimesis*), arguing that it operates "away from the truth" and appeals to the irrational part of the soul (Plato, *Republic*, 603b). Building on this, critics have been arguing that film "lacks the explicitness to formulate and defend the precise claims characteristic of philosophical writing" (Wartenberg 2007, 16). In other words, cinema cannot construct explicit philosophical arguments.

Bruce Russell is one of the leading proponents of this objection, and he examines this issue through Louis Malle's *Le Feu Follet*. Russell addresses Granger's claim that films can present philosophical arguments in analysing the film. In the movie, the protagonist, Alain Leroy, contemplates the meaninglessness of life, engaging with the fundamental philosophical question of whether life is worth living (Russell 2008, 1). Granger contends that even if the film's argument is insufficient to "yield knowledge, or even justified belief, of its conclusion," it still presents a philosophical position (Ibid., 17). This assumes a distinction between films making philosophical *claims* and films making philosophical *arguments*. While a film may suggest or illustrate a position, critics like Russell argue that it cannot *explicitly* construct or defend an argument in the rigorous manner characteristic of philosophical discourse. Russell challenges this view by highlighting the inherent limitations of cinematic argumentation. He argues:

> Even if it is possible for a narrative film to offer an implicit argument, I do not think that such an argument would do more than merely raise a philosophical question. (...) Too many crucial premises remain unstated and undefended for it to be said that I do more than raise a philosophical question (Ibid., 19).

Russell's position stems from his understanding that merely raising questions does not constitute philosophy. For him, the essence of philosophy lies in explicit argumentation that justifies beliefs about general claims. This leads to his argument connecting the explicitness objection with broader philosophical concerns:

1) Philosophy involves justifying beliefs about general philosophical claims.
2) Since answers to philosophical questions are not obvious, explicit argumentation is necessary for this justification.
3) Films do not provide explicit argumentation.

4) Thus, films cannot justify beliefs about general philosophical claims (Ibid., 20).
5) Therefore, films cannot philosophise.

Murray Smith further develops this argument by highlighting the institutional differences between art and philosophy. Smith argues that film cannot philosophise due to its inherent ambiguity as an art form. While ambiguity is valued in art, philosophy considers it a defect. Smith emphasises that works of art are valued for their "complexity, ingenuity, inventiveness, density, ambiguity, and profundity" (Smith 2006, 40), whereas philosophy is valued for its epistemic clarity.

Smith proposes three specific norms for philosophy:

a) Rational warrant: This refers to evaluating the reasons behind our beliefs and practices, focusing on whether they are rational or irrational. It serves as a key tool for assessing the logic of our beliefs. In philosophical discussions outlined by Smith, we aim to determine the extent of rational warrant supporting our beliefs. Whenever possible, we should seek to revise those beliefs and replace them with ones that have a stronger rational warrant.
b) Empirical adequacy: This is crucial in evaluating a philosophical thesis, according to Smith. Philosophers must examine the evidence to support or refute specific philosophical hypotheses and adapt them based on empirical findings.
c) Reflective maturity: Inspired by John Rawls' concept of "reflective equilibrium," Smith suggests that our philosophical conclusions should arise from a thorough reflection on our beliefs and assumptions from multiple perspectives (Smith 2016, 187-89).

As a result, Smith argues that philosophy is more demanding than science and significantly more challenging than art in terms of the criteria of rationality, empiricism, and reflexivity. This contrasts with art's primary objective, which Smith describes as delivering "an adventure in perception, cognition, and emotion. Artworks embody an array of values—aesthetic, moral, epistemic—that interpenetrate one another" (Smith 2006, 185).

The distinction between art and philosophy is also illustrated through Paisley Livingston's "propriety objection," which draws from Hegelian philosophy. Hegel asserts that "the poetic work of art has no aim other than the production and enjoyment of beauty; in its case, aim and achievement lie directly in the work itself (...) the artistic activity is not a means to a result falling outside itself but an end which in its accomplishment directly closes together with itself" (Hegel 1975, 992). Livingston, thus, argues that art has intrinsic value and should not be reduced to something cognitive (Livingston 2009, 39-41).

The ambiguity of art "arises from the concreteness and particularity of art," whereas philosophy is valued for its "abstract, conceptual character." As Smith notes, "[f]ew criticisms are more apt to strike terror into the heart of the philosopher than the assertion that such-and-such a proposition is 'ambiguous'" (Ibid., 40), while in the world of art, ambiguity is more apt to be used as a term of praise.

To support this point, Smith cites the example of Bernard Williams' thought experiment about personal identity and the film *All of Me* (1984) by Carl Reiner.

In his thought experiment about personal identity, Williams imagines a man called Charles who undergoes a character transformation. After waking up, Charles claims to have memories of events he had not recalled previously, while he also loses his own memories. These new memories correspond to the life of Guy Fawkes. If Charles has Fawkes' memories, has Fawkes returned to Charles' body? (Williams 1957). This scenario is supposed to question what the connection between memory and personhood is, challenging assumptions about what makes someone the same person over time.

Smith argues that while Williams' thought experiment has a primarily epistemic role, *All of Me* is primarily humorous. Smith argues that film and philosophy have different priorities, and the difference of objectives in Williams' thought experiment and *All of Me* shows that. This movie follows the story of a lawyer named Roger Cobb, who is assigned by his boss to handle the legal matters of a wealthy client named Edwina Cutwater. Edwina is on the brink of death and seeks the help of a guru to transfer her mind into the body of Terry Hoskins. Roger's task is to deliver the will to Terry so that she can inherit Edwina's fortune after the mind transfer. However, an unexpected incident occurs during the process, causing Edwina's mind to end up in Roger's body. This results in a comical situation where both minds inhabit the same body.

Smith argues that while the film raises questions about personal identity, its primary purpose is to provide entertainment and humour. As Smith notes, this reflects a sentiment often attributed to Samuel Goldwyn, who stated in various ways the idea that "If you want to send a message, use a telegram" Pictures, in this view, are for entertainment, not philosophical messaging.

Smith emphasises that "the vocation of art is to deliver an adventure in perception, cognition, and emotion. Artworks embody an array of values – aesthetic, moral, epistemic – that interpenetrate one another" (Smith 2016, 185).

Smith refers to a list proposed by Paisley Livingston, which includes both philosophical works and films, in order to further his argument:

Edmund Husserl, *Logical Investigations*
Ludwig Wittgenstein, *Philosophical Investigations*
W. V. O. Quine, *Word and Object*
John Rawls, *A Theory of Justice*
Thomas Kuhn, *The Structure of Scientific Revolutions*
Saul Kripke, *Naming and Necessity*
Frank Capra, *It Happened One Night*
Laurence and Andrew Wachowski, *The Matrix* (Livingston 208, 590).

This list includes two films alongside several philosophical works. According to Livingston, we find this strange because we have a tendency not to view films as philosophical works. This perspective is likely based on reasons that should not be arbitrary. Smith argues that the titles of films are strange because filmmakers are not typically regarded as philosophers. He notes that the distinction between "art" and "philosophy" is not merely a historical coincidence but a significant difference. This distinction largely arises from the fact that art and philosophy serve different purposes.

He refers to Cleanth Brooks' "The Heresy of Paraphrase," which states that perhaps we should not force "the statement (of a poem – by extension an artwork) into an unreal competition with science or philosophy or theology" (Brooks, cited in Smith 2006, 185). In line with this, Sinnerbrink resists the framing of film as a failed version of philosophy, as we have seen, and instead argues that film engages in a "distinctively cinematic kind of thinking" (Sinnerbrink 2011, 7). Rather than competing with philosophy, film contributes to philosophy by transforming how we approach concepts through aesthetic experience. As such, Sinnerbrink pushes back against the reduction of film to philosophical purposes, advocating for a model in which art and philosophy are mutually transformative rather than hierarchically ordered – the film-philosophy model.

To show how movies are limited for doing philosophy, Smith gives the example of Ridley Scott's *Blade Runner* (1982), stating that its power and value

> derive not from the fact that it articulates what we recognize as philosophical ideas, and still less from its establishing the truth or plausibility – in a way meeting the philosophical standards I've discussed – of these ideas. As a futuristic fictional narrative, the film offers

very little in the way of systematic empirical support for the philosophical ideas it represents, nor does it subject these ideas to searching rational tests, and thus it cannot deliver reflective maturity" (Smith 2016, 193).

In addition to this film, Smith also uses the example of Joshua Oppenheimer's *The Act of Killing* (2012), arguing that this film does not provide "the evidence presented from numerous angles, offering counterexamples and counterarguments, seeking to establish reflective maturity in the philosophical sense described earlier" (Ibid., 194).

In short, Smith's differentiation between art and philosophy lies in their different functions and the way these two are evaluated: works of art are not evaluated by their philosophical content but for their own sake (Ibid., 196).

Finally, Smith affirms that his argument is not eliminativist but deflationary. For him, art may have cognitive value, but not in the same way as philosophy (Ibid., 197). However, Smith ends up accepting that, because works of art may have cognitive value, philosophers may appropriate them, and that there may be an "area of overlap between the domains of art and philosophy" (Ibid., 197).

2.3 The generality objection

Related to the explicitness objection, there is also the generality objection, made by Bruce Russell. The generality objection is rooted in the idea that philosophy seeks to establish universal or general truths, while films, especially fictional narrative films, present particular and often fictional scenarios. As Russell argues, films can vividly introduce philosophical problems and even solve some by showing what is possible. However, they cannot demonstrate what is probable or, in some cases, even what is possible (Russell 2000, 163). This limitation comes from the fact that films produce specific narratives that cannot serve as evidence for general philosophical claims.

This objection is related to what Noël Carroll calls the *no-evidence argument* made by Monroe Beardsley. Beardsley's critique focuses on the epistemic limitations of fiction, arguing that fictional works cannot provide the kind of evidence required to support philosophical hypotheses. In his analysis of *Baby Doll* (1956), directed by Elia Kazan, Beardsley argues that the film "does not present a single authenticated case history (being fiction), and therefore it certainly is not strong or direct *evidence* for the hypothesis [which the film claims to present]" (Beardsley 1958, 380). In

other words, while the film may suggest philosophical ideas, it cannot provide the necessary justification for these ideas because it is a fictional narrative.

Here emerges a twofold challenge to film as philosophy. First, even if a film presents a particular case, its fictional nature prevents it from providing the empirical evidence required for philosophical generalisation. Second, the fictional status of these cases makes it difficult to infer any evidence from them. The generality objection complements the explicitness objection: film is not able to explicitly formulate and defend philosophical claims, and it does not have the capacity to generalise from non-actual, particular instances.

Bruce Russell's critique of the philosophical potential of films is elaborated through his analysis of specific films, particularly *Crimes and Misdemeanors* (Woody Allen, 1989) and *A Simple Plan* (Sam Raimi, 1998). *Crimes and Misdemeanors centres on Judah Rosenthal,* who orchestrates the murder of his mistress, Dolores, to prevent her from revealing their affair and ruining his reputation. Initially, Judah is consumed by guilt and fear, but as time passes, he finds that his life improves, and he can move on without facing any significant consequences for his actions.

Russell argues that *Crimes and Misdemeanors* serves as a counterexample to the philosophical thesis that wrongdoing necessarily leads to unhappiness or conflicts with one's self-interest. The film illustrates a possible outcome of a crime – Judah's life improves after the murder – but it does not establish a general truth about the consequences of wrongdoing. As Russell notes, "no one can establish that something holds in all possible worlds by presenting an example or two of a possible world depicted in film" (Russell 2000, 163). While the film raises philosophical questions about morality, guilt, and justice, it cannot provide the explicit argumentation or evidence required to support a universal philosophical thesis.

Russell contrasts *Crimes and Misdemeanors* with *A Simple Plan* (Sam Raimi, 1998), which he argues presents a more probable outcome of criminal behaviour. In *A Simple Plan*, three characters, Hank Mitchell, his brother Jacob, and their friend Lou Chambers, find a crashed plane containing millions of dollars. They decide to keep the money, but their decision leads to a series of tragic events, including betrayal, violence, and death. Hank, the protagonist, becomes unhappy and morally compromised as a result of his actions. For Russell, *A Simple Plan* shows that crime does not pay. However, even in this case, Russell maintains that the film's narrative is limited to a particular scenario and cannot establish a general philosophical truth.

Now, it may be appropriate to question whether Russell is not at odds with his own criticism. For the philosopher, one or a few cases or examples are not sufficient to make a general or universal statement. Russell makes a general statement: *no* film makes relevant philosophy. However, the philosopher makes this claim based on two examples he provided (*Crimes and Misdemeanors* and *A Simple Plan*). According to Russell, these examples show how films can only raise questions and function as counterexamples to philosophical theses, and nothing more. But in doing so, does not the philosopher contradict his idea that examples cannot lead to a general thesis? Thus, by his argument, his reasoning for the general thesis that films only raise questions and counter-exemplify philosophical theses is invalid.

2.4 The appeal to thought experiments

A common tactic used by defenders of literary cognitivism and FAP is to appeal to the use of thought experiments in philosophy. Wartenberg argues that the persistent idea that philosophy always seeks general and universal truths may be wrong. He argues that philosophy can be made through film, similar to how film can serve as a medium for mathematics or history. If philosophy is as abstract and universal as mathematics, defending FAP becomes more challenging, since "the sort of intellectual operations that are characteristic of mathematics are so different from the ways in which film operates that it is hard to even begin to speculate on how fiction film might be able to count as screening of mathematical knowledge" (Wartenberg 2007, 22).

Philosophy may share some similarities with history. In historical research, one common method is the use of narratives, stories told by individuals about specific events. For example, Wartenberg highlights the 1977 documentary *With Babies and Banners* by Gary Lorraine, which offers a feminist reinterpretation of the strike against General Motors in Michigan that took place from 1936 to 1937. This suggests that if philosophy, like history, utilises narratives as a means of producing knowledge, there could be a form of philosophising that draws upon specific cases.

In philosophy, narratives and thought experiments play various roles. Bruce Russell acknowledges that films can serve as counterexamples to philosophical theses; however, he does not seem to regard this as a significant philosophical contribution. This is particularly important considering that counterexamples are thought experiments.

As Noël Carroll notes thought experiments are:

> devices (...) frequently employed by philosophers to defend and/or to motivate their claims, moral and otherwise. Thus, if these strategies are acceptable forms of knowledge production in philosophy and if literature contains comparable structures, then if philosophy conducted by means of thought experiments is an adequate source of knowledge and education, then so should literature be (Carroll 2002, 7).

Catherine Elgin, too, argues that "just as thought experiments are fictions in science, works of fiction are thought experiments in art" (Elgin 1993, 25). This helps explain how fictional narratives, being particular cases, can convey general philosophical ideas. For Elgin, scientific thought experiments and fictional narratives work by exemplifying particular features. While a thought experiment or fictional narrative presents a specific scenario, its cognitive value lies not in establishing empirical facts about that scenario but in illuminating patterns and relationships that can be recognised in other contexts. This helps address Russell's concern about deriving general conclusions from particular cases.

Furthermore, Elgin argues that the success of thought experiments and fictional narratives depends on the accuracy and adequacy of their background assumptions rather than if they are verifiable or not. She notes, for example for a scientific thought experiment on resistance and temperature, "the simulation discloses that at absolute zero, electrical resistance disappears. It does not, of course, demonstrate that absolute zero ever in fact is reached, or that resistance ever in fact disappears. But by revealing what would happen in the limit, it enhances our understanding of the connection between resistance and temperature" (Elgin 1993, 24). Similarly, a film does not have to present a real situation to contribute to our understanding of philosophical concepts.

This suggests that the particularity of fictional narratives, rather than being a limitation, is essential to their philosophical function. The key is not whether the specific scenario actually occurs but whether it mobilises our understanding.

Wartenberg analyses "cinematic thought experiments" in detail. As previously mentioned, he analyses *The Third Man* and also *The Matrix* (1999) by the Wachowski sisters. From his perspective, *The Matrix* illustrates Descartes' sceptical claim that we can have the same experiences we do today without the existence of an external world. These experiences and perceptions could potentially be orchestrated or manipulated by an entity, such as an evil demon. Descartes proposed the possibility that a powerful demon might be deceiving him about everything he experiences. He contemplates that the beliefs he holds to be the most certain about the external world, mathematics or even his own body, might be illusions implanted by that deceiver. In

considering this possibility, Descartes temporarily adopts a radical doubt: “I will suppose then, that everything I see is spurious. I will believe that my memory tells me lies, and that none of the things that ir reports ever happened. I have no senses. Body, shape, extension, movement and place are chimeras. So what remains true?” (Descartes 1996, 16). Wartenberg terms this idea the “deception hypothesis,” which is central to the narrative of *The Matrix*.

When watching this film, viewers are confronted with the question of whether they might be living in an illusion or a simulation. The narrative centres around the main character Neo, who discovers that his reality is a simulation known as the “Matrix,” created by machines that have dominated the world. These machines use living human bodies for energy supply, and create their experiences of the world in order to keep them sleeping and not resisting it.

Rather than detailing the entire film, let us focus on the most relevant moments to the argument being made. One important scene occurs with a mirror, where Neo sees himself reflected in a flexible surface. During this moment, Morpheus, the leader of the human rebels fighting to overthrow the Matrix and free humanity, questions Neo whether he has ever experienced a dream that felt real. He then asks how Neo can distinguish dreams from reality. As their dialogue unfolds, Neo touches the mirror, which begins to move through his arm and body, leading to a cardiac arrest. This moment is pivotal, as the screen then turns black while the rebels locate and rescue Neo’s physical body.







Figure 6: Still images from *The Matrix* (directed by Lana and Lily Wachowski, Warner Bros., 1993)

According to Wartenberg, one of the most original elements of this film as a thought experiment in deception and simulation is that its narrative allows viewers to follow and share Neo's own experiences, which are clearly presented as merely creations of the Matrix. That is when we watch this film, "we initially believe in the (fictional) reality of the world that is depicted by the film. As Neo comes to see that the world he had believed to be real was only a computer simulation, *so do we*" (Wartenberg 2007, 150). What happens, then, is that this narrative leads the viewer to question whether what is shown to her on the screen cannot be her own reality. Another important scene is when Neo has a nightmare about being interrogated by the Matrix agents he wakes up from. In this case, the film "relied on a technical possibility (...) to frame a sequence in such a way that what the audience took to be part of the screened world was really just a subjective experience of one of that world's characters" (Ibid.). This film positions the viewer in an epistemic situation similar to that of Neo, ultimately leading her to consider the deception hypothesis.

While Wartenberg treats *The Matrix* as a thought experiment regarding deception or simulation, he acknowledges two key points: 1) certain specific scenes in the film support this interpretation, but not the entire film does, and 2) the film does not significantly advance philosophical discourse, as it is not particularly original. Nevertheless, one can argue that the film

induces sceptical doubts in viewers, much like a philosophical text would. It may be even more effective, as it manages to mislead the audience within the narrative; viewers only discover the reality of the Matrix alongside Neo.

This film, like many other works of fiction, shares a common trait with philosophical thought experiments: it encourages the audience to imagine or see fictional scenarios. Consequently, *The Matrix* prompts viewers to reach conclusions similar to those derived from Descartes' thought experiments. Therefore, it can be considered a thought experiment that explores the possibility of deception or simulation.

Wartenberg identifies the same specific type of thought experiment that can be found in films that Russell does: the counterexample. Philosophical theses are often general or universal. A counterexample, however, whether real or imagined, challenges the validity of such general claims. One example he discusses is found in Plato's *Republic*, which challenges the idea that it is always just to speak the truth and return what does not belong to us. In response to Cephalus' claim, Socrates presents a counterexample: he imagines a scenario in which a friend has loaned us a weapon but later demands its return while in a state of madness. In this situation, Socrates argues that it would not be just to return the weapon or to tell the truth, questioning the concept of absolute justice.

According to Wartenberg, a counterexample consists of two key components: a) the specific counterexample itself, which can be seen as a thought experiment that highlights a particular instance; and b) the reasoning that demonstrates why this scenario serves as a counterexample to a general philosophical thesis, relying on deductive reasoning. This reasoning occurs in the following manner:

1) we start with a general proposition of the form "all x are y";

2) someone provides a counterexample indicating "there exists a case of an x that is not y";

3) this leads to the conclusion that the initial proposition (1) is false.

With this analysis of what constitutes an effective thought experiment as a counterexample, he presents a film that can be considered a counterexample to a philosophical thesis. This film is *Eternal Sunshine of the Spotless Mind* (2004), directed by Michel Gondry, which appears to challenge the principles of utilitarianism.

The film revolves around the characters Joel Barish and Clementine Kruczynski. These two characters, who meet by chance, discover that they have erased all memories of their past painful romantic relationship. This memory erasure is made possible by a technology developed by Dr. Howard Mierzwiak, which allows individuals to selectively eliminate memories that cause them distress.

Wartenberg argues that this technology could be justified on utilitarian grounds. A utilitarian might suggest that it enables people to remove memories that cause suffering, thereby improving their quality of life. In the film, the characters seem to embrace this perspective exactly. When Joel learns that Clementine has erased her memories of him, he decides to do the same. During Joel's preparations for the memory erasure, Mary, an assistant, presents a utilitarian argument for the procedure. Her argument is straightforward: if painful romantic memories cause emotional suffering, and the technology can remove that suffering, then erasing those memories is morally justified because it increases the individual's overall happiness and quality of life. From this perspective, the reduction of suffering outweighs the loss of the memories themselves.

The film also shows the issues associated with the idea of erasing memories in several ways. At the end of the story, Joel, Clementine, and even Mary, who had an affair with Howard, find themselves in the same problematic relationships they had before eliminating their memories. This repetition is what Wartenberg refers to as the "recidivism problem," (Ibid., 88), the tendency to repeat behaviours that led to painful memories in the first place. The film suggests that this pattern is a consequence of the technology used to erase memories.

The film makes a statement about the important role that all memories, including negative ones, play in our lives. Wartenberg discusses the "educative function of memory in relation to desire" (Ibid., 89), emphasising how our memories are crucial in deciding whether to fulfil a certain desire, especially when past fulfilment has caused us harm. The film insists that erasing painful memories is problematic. The main issue with this technology is that by erasing painful memories, we also lose the learning processes that depend on those experiences. Additionally, another idea emerges in the film. As Joel undergoes the memory-erasing process, he consciously realises that his painful memories of his breakup with Clem are intertwined with happy memories from their relationship. However, he is unable to communicate his desire to stop the procedure because he is asleep and under the machine's control. This scene shows the film's concern that surrendering "control over our minds to others" leaves us powerless to regain that control. In this context, the

film shows a harmful heteronomy between the individual whose memories are erased, and for those who participate in the procedure (Ibid., 90).

Overall, this film can be viewed as a thought experiment that challenges certain forms of utilitarianism, particularly act utilitarianism focused on maximising pleasure and minimising pain in the present moment, by presenting the idea that the memories that cause suffering can also be beneficial for our lives, particularly in the learning process. The film specifically challenges the assumption that immediate happiness should always be prioritised over personal development. However, while this may not definitively refute Mill's version of utilitarianism, which already values higher pleasures and personal development, it does highlight tensions within utilitarian thinking about how to weigh immediate well-being against long-term flourishing and self-improvement properly.

While there is debate about whether thought experiments equate to arguments, an issue to explore later, the essential point is that this film serves as a thought experiment and is a useful tool in philosophy.

2.5 The role of documentary in film as philosophy

An interesting note that can be added to previous objections is that many philosophers who remain sceptical about fictional films readily acknowledge that documentaries can philosophise. This reveals some assumptions about the relationship between fictional narrative and philosophical content that deserve some analysis, particularly how the fiction/nonfiction distinction shapes these objections.

Despite his general scepticism toward FAP, Smith acknowledges that certain documentary works can make philosophical contributions. He notes that *The Act of Killing* possesses some cognitive value, though he maintains that it falls short of meeting the standards he set for philosophical work. His critique centres on ambiguity and a lack of explicit argumentation of fictional films, as previously seen. Yet, these same concerns appear to diminish or even dissolve when applied to documentaries, suggesting that the fictional status of narrative films is itself viewed as the main problem.

Similarly, Wartenberg seems to draw a distinction between fictional films and documentaries in their philosophical potential when he explicitly recognises works like *Ways of Seeing* as legitimate works of philosophy. The implication is that the nonfictional status of

documentaries would enable them to make direct philosophical claims about reality in ways that fictional narratives cannot.

The generality objection articulated by Russell exemplifies this concern. Russell argues that films cannot demonstrate general philosophical truths because they present particular, often fictional scenarios that cannot serve as evidence for universal claims. This objection gets much of its force from the assumption that fictional narratives, by definition, cannot provide the kind of evidence required for philosophical generalisation. But when the same sceptics encounter documentaries, this objection seems to lose part of its appeal. The nonfictional status of documentary seems to restore the value of evidence that fiction seemingly does not possess.

However, this distinction between fictional and documentary cinema may be less obvious than these objections may assume. As Vítor Moura demonstrates in his analysis of documentary identification, the boundaries between fictional and nonfictional film are more fluid than they may appear. (Moura 2019). Documentaries are not simply transparent windows onto reality but are constructed representations. Moura notes that "all representation is essentially fictional, and film necessarily involves a number of stylistic, narrative and perspectival choices", and this should make us sceptical of traditional realist views (Ibid., 604-605). If this is correct, then the objections based on fictionality would apply equally to documentaries. This complexity, identified by Moura, extends to how viewers engage with documentary films. He notes that documentaries involve a tension between what is being "told" and what is being "shown" (Ibid., 607), creating what he calls a "falsificatory assessment" where viewers ask, "might it be lying?" (Ibid., 624). This epistemic uncertainty suggests that documentaries face many of the same challenges that we can attribute to fictional films. If documentaries can deceive, manipulate, and present biased perspectives, which they can[9], then their nonfictional status provides less epistemic security than we may assume.

Moreover, Moura's analysis reveals that the fiction /nonfiction distinction may be more a matter of viewing attitude than inherent properties. He argues that the difference between

[9] A striking example is Animal Planet's *Mermaids: The Body Found* (2011). It is important to note that Animal Planet is known for its documentaries on animal life. This mockumentary presented entirely fictional "evidence" of mermaid discoveries using documentary conventions so effectively (such as what appeared to be real footage and interviews with "scientists") that viewers genuinely believed they were watching factual content. The program became a pop culture phenomenon, with many audience members convinced of the reality of the "discoveries" and sharing their beliefs across social media platforms. This widespread cultural impact demonstrates how documentary form can make even obviously fictional content appear authentically factual.

documentary and fiction film is determined by how the audience recognises the author's intentions (Ibid., 618). This account may undermine objections that rely on ontological distinctions between fictional and nonfictional content.

The implications for philosophical thought experiments may be considerable. If the fiction/nonfiction distinction is less stable than critics assume, then objections based on fictionality lose much of their force. The generality objection, for instance, assumes that fictional scenarios cannot provide evidence for philosophical claims. But if documentaries themselves involve fictional elements and interpretation, then the relevant question is not whether content is fictional or nonfictional, but whether it effectively philosophises.

Furthermore, if we want to accept that documentaries can philosophise, or if we are at least more inclined to believe so, we must acknowledge that they often do so through precisely the kinds of narrative, visual, and rhetorical techniques that characterise film more broadly. For example, *The Act of Killing* can be viewed as philosophical because of its formal strategies like role-playing and reconstructions. These techniques are not fundamentally different from those in fictional films. Thus, we can suggest that it is a cinematic technique that enables philosophical work, not factual or nonfactual content.

In this regard, objections based on formal or epistemic concerns may actually rest on unstable assumptions about the nature of fictional and nonfictional representation. Thought experiments, whether textual or cinematic, are themselves fictional scenarios designed to illuminate philosophical concepts. The fact that they are imaginary rather than actual does not exclude them from epistemic value, any more than the constructed nature of documentaries or fictional films does. Thus, rather than asking whether films are fictional or nonfictional, we should ask whether particular films employ their cinematic techniques in philosophical ways, as we shall see.

This chapter examined some of the main objections to film as philosophy: the banality objection, the explicitness objection and the generality objection. The banality objection, primarily articulated by Stolnitz, argues that artistic works offer little cognitive value, merely reiterating trivial truths. Carroll challenges this view, suggesting that what appears to be banal for experts may be revelatory to broader audiences. The explicitness objection made by Kivy and Smith, contends that films cannot construct explicit philosophical arguments. This critique emphasises institutional differences between art and philosophy. The closely related generality objection introduced by

Bruce Russell suggests that films are limited by their particular narratives and cannot establish universal truths. This objection challenges the epistemological potential of fictional narratives.

In response to these objections, I explored the appeal to thought experiments. Wartenberg and Carroll draw parallels between cinematic narratives and philosophical thought experiments. Using examples such as *The Matrix* and *Eternal Sunshine of the Spotless Mind*, they tried to show how films can function as philosophical thought experiments.

An examination of how these objections apply differently to documentary versus fictional films reveals important assumptions underlying the criticism of film as philosophy. While there is scepticism about fictional films' philosophical potential, there is more readiness to acknowledge that documentaries can philosophise. However, this distinction rests on unstable assumptions about the fiction/nonfiction boundary. We saw that documentaries are themselves constructed representations. I suggest that the relevant question is not whether content is fictional or factual, but whether films employ their cinematic techniques in philosophically productive ways.

These discussions establish the foundation for examining what thought experiments are, their role in philosophy, and exactly how films might function as effective philosophical thought experiments.

3. Thought experiments in science and philosophy

Thought experiments are not exclusive to philosophy. They are used in several other disciplines, including science, as they explore imaginary scenarios. Tamar Gendler defines the thought experiment as "a process of reasoning carried out within the context of a well-articulated imaginary scenario in order to answer a specific question about a non-imaginary situation" (Gendler 2000, x). For Gendler, a thought experiment is to "reason about an imaginary scenario with the aim of confirming or disconfirming some hypothesis or theory" (Gendler 2004, 1154). Roy Sorensen adds another important aspect: "Thought experiments must be presented with the public aim of answering or raising the experimental question without the benefit of execution" (Sorensen 1992, 196).

We can trace thought experiments to pre-Socratic philosophy, and ever since, they have been a constant presence in philosophy. Lichtenberg is referenced for his reflections on the nature of thought experiments, noting how thought experiments can be used to explore and illustrate ideas, suggesting that they can serve as means to justify beliefs and expand our understanding of concepts (Fehige and Stuart 2014, 8). Hans Christian Ørsted is another significant contributor to the investigation of thought experiments, especially in the philosophy of science. He has been attributed the coining of the term "thought experiment" ("Tankeexperiment") in 1811 (Ibid., 1), as he viewed it as a "creative method" that helps scientists conceptualise and understand phenomena that might not be able to be observed or verified. These ideas influenced other philosophers and scientists, including Ernst Mach.

Mach argued that thought experiments are tools for exploring concepts and testing hypotheses without physical experimentation. This allows scientists and philosophers to engage with hypotheses that may be difficult or impossible to test empirically (Mach 1976).

Thomas Kuhn also analysed thought experiments in science, arguing that they serve several functions, including clarifying scientific concepts by revealing contradictions and exploring the implications of theories in situations that may not be tested (Kuhn 1977(. Thought experiments can reveal the limitations of current knowledge by challenging current assumptions. For example, Einstein's *train* thought experiments challenge our understanding of time, and the *Heinsenberg Microscope* shows the limitations of empirical testing in quantum mechanics.

Einstein describes a train moving at constant speed past a platform. Lightning strikes both ends of the train simultaneously for an observer on the platform. An observer on a train sees strikes

occur in different times. For the train passenger, the strike at the front reaches them before the strike at the back. The platform observer sees equal light travel time from both strikes. Events co-occurring for one observer occur at different times for another moving relative to the first. This reveals that time measurements depend on the observer's motion, and why light speed remains constant while time dilates, showing how special relativity resolves this.

The Heinsenberg Microscope thought experiment shows the limitations in measuring quantum particles. Light must bounce off an electron to be detected when locating its position. This creates a problem: using visible light gives a blurry image because its wavelength is much larger than the electron. Using gamma rays with wavelengths closer to the electron's size causes photons to collide with it, changing its position. Heisenberg claimed that this disturbance cause uncertainty, which Bohr criticised. Quantum physicists thought that this uncertainty came from measurement limitations and not because it was an intrinsic property. However, the uncertainty principle is currently accepted, recognising that indeterminacy is inherent to quantum systems.

These examples of thought experiments help us explore and challenge scientific principles (Kuhn 1977, 240-65).

This chapter treats these scientific and philosophical thought experiments together for several reasons. First, both types of thought experiments share structural features: they involve counterfactual reasoning, imaginative scenarios, and claims to generate new insights through mental manipulation rather than empirical investigation. Many of the core epistemological puzzles they raise are similar: how can mere thinking produce knowledge? What is the relationship between intuitive responses and justified belief? How do we account for their apparent epistemic power? As shall be seen, several of the theoretical frameworks for understanding thought experiments apply equally well to both scientific and philosophical cases. By examining both together, we can develop a more comprehensive understanding of thought experiments as a general cognitive and epistemic tool.

3.1 The paradox of thought experiments: the Platonic versus the Argument view

Contemporary debates about thought experiments revolve around how imaginary scenarios that happen only in our minds can tell us anything new about scientific or philosophical matters. Some of the questions we may ask about thought experiments are: do thought experiments provide us with any new knowledge? What kind of epistemic value do they have, if any? How can we distinguish

good from bad thought experiments? The main question about thought experiments is a famous paradox, or puzzle: "Thought experiments often have novel empirical import even though they are conducted entirely inside one's head" (Horowitz and Massey 1991, 1). This puzzle led to a divide between two main positions: on the one side, Platonists like James Brown contend that thought experiments can give us direct access to abstract laws of nature; on the other side, empiricists like John Norton argue that thought experiments are just arguments in disguise. As will be seen, there are also alternatives to these two opposing views.

3.1.1 The Platonic view

James Brown argues that thought experiments can sometimes lead to *a priori* knowledge of nature, revealing truths about the world without relying on new empirical observations. Brown's position is based on his analysis of what he calls "Platonic thought experiments", which are rare cases that both destroy an old theory and generate a new one without any empirical input:

> A *Platonic thought experiment* is a single thought experiment which destroys an old or existing theory and simultaneously generates a new one; it is *a priori* in that it is not based on new empirical evidence, nor is it merely logically derived from old data; and it is an advance in that the resulting theory is better than the predecessor theory (Brown 2011, 99).

So, Brown argues that there are two main types of thought experiments: destructive and constructive, which can be subdivided into direct, conjectural and mediative.

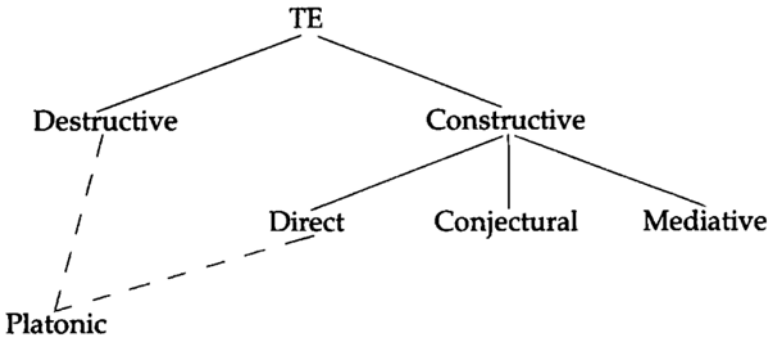


Figure 6: Brown's scheme for types of thought experiments (Brown 1991, 33).

A destructive thought experiment is a "picturesque *reductio ad absurdum"* (Ibid., 33), aiming to destroy or identify problems in a theory. They do so by showing internal inconsistencies

in a theory (for example, how Galileo showed fault in Aristotle's theory of free fall), or showing how a theory can conflict with common sense (for example, how Schrödinger showed that quantum theory goes against itself and common sense) (Ibid., 35).

Galileo challenged Aristotelian physics in his falling bodies thought experiment. He thought about two objects with different weights being dropped from a tower. Aristotle claimed that a heavier object falls faster than a lighter one. Galileo asked what happens if objects connect during a fall. The combined object weighs more than either part alone. This suggests that a combined object should fall faster than a heavy object alone. However, the lighter part should slow the heavier part. This would create a contradiction. Galileo resolved this paradox by concluding that all objects fall at the same rate, regardless of mass. It would not be the weight, but the air resistance that would cause differences in fall rates. This reveals that laws of motion apply equally to objects.

Schrödinger's Cat illustrates the paradoxical aspects of quantum mechanics when applied to everyday objects. In this thought experiment, a cat is placed in a sealed box with a radioactive atom, a Geiger counter, and a vial of poison. If the atom decays, the counter detects it, triggering the release of the poison, killing the cat. According to quantum mechanics, until observation occurs, the atom exists in decayed and undecayed states in superposition, making the cat alive and dead simultaneously. This shows the measurement problem in quantum theory and questions whether quantum indeterminacy can extend beyond microscopic systems.

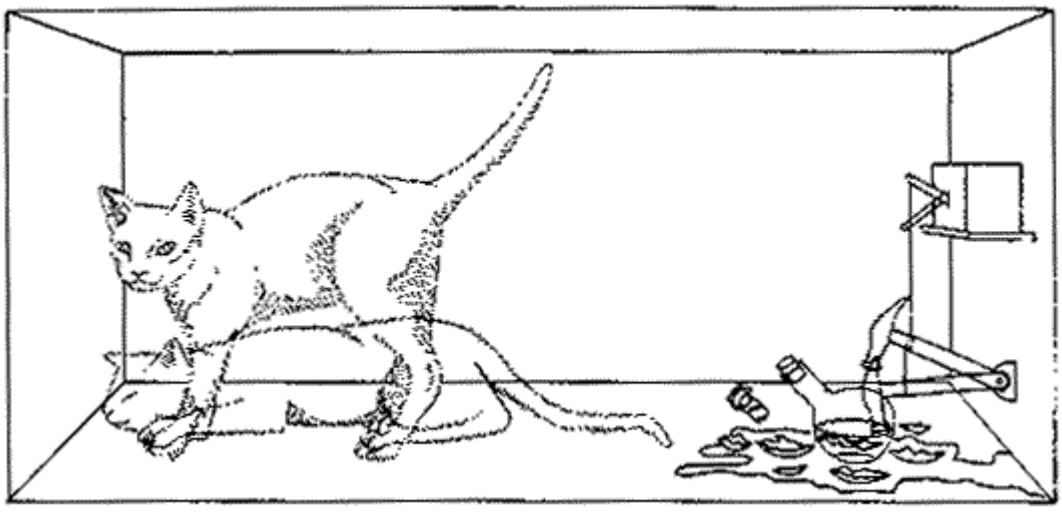


Figure 7: "Schrödinger's Cat," by Erwin Schrödinger, 1935. Source: *University of Toronto Physics Department*, accessed April 22, 2025.

Constructive thought experiments, on the other hand, aim to establish a scenario that evidences a theory: "The burden of any constructive thought experiment is establishing (in the imagination) the thought-

experimental phenomenon. This phenomenon then acts as fairly conclusive evidence for some theory" (Ibid., 43).

As seen earlier, there are three types of constructive thought experiments. The *mediative thought experiment* helps reach a conclusion based on a specific theory, and it can do so by, for example, illustrating some counterintuitive aspect of a theory. An example of this that Brown refers to is Maxwell's demon (Ibid., 35-38). Maxwell's demon, proposed by James Clerk Maxwell, challenges the second law of thermodynamics. In this scenario, a hypothetical being (the demon) controls a small door between two gas-filled chambers. By selectively allowing faster molecules to pass in one direction and slower ones in the opposite direction, the demon could create a temperature difference between the chambers without expending work, seemingly decreasing entropy. This apparent violation caused examination of the relationship between information and thermodynamics, leading to the conclusion that the demon's process itself would require energy, preserving the second law of thermodynamics (natural processes tend to increase disorder, or entropy, in an isolated system).

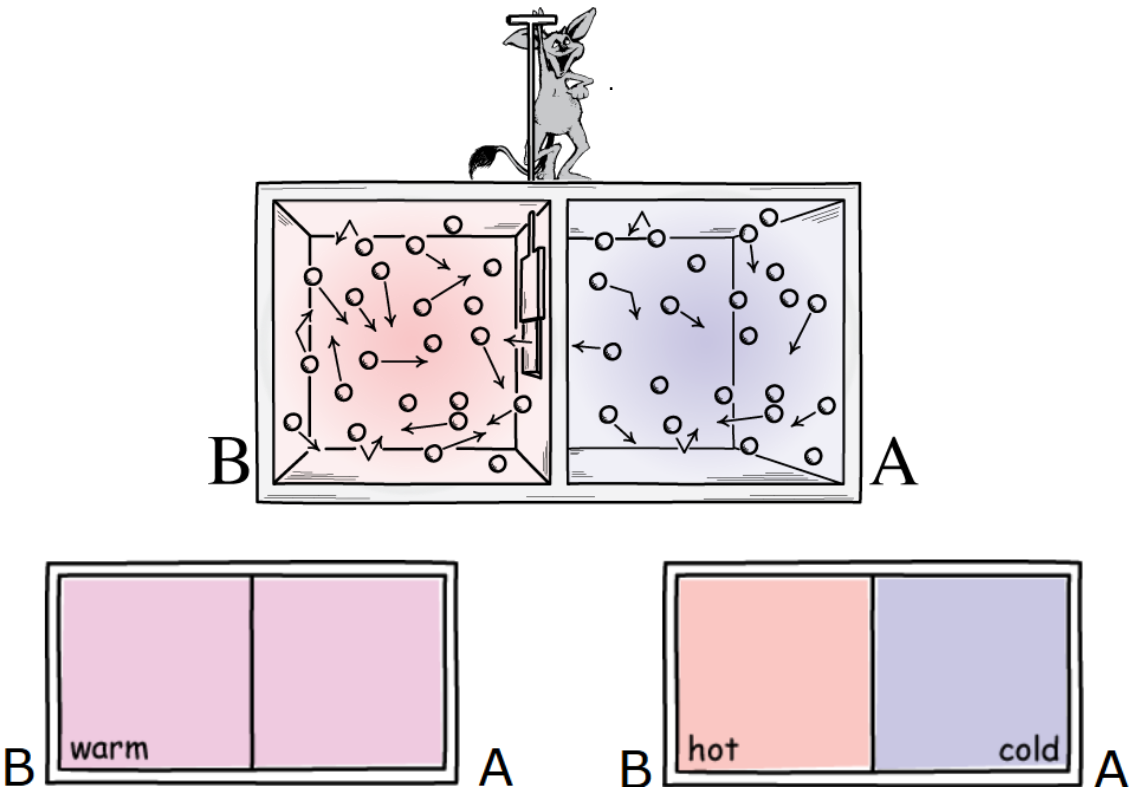


Figure 8: Illustration from John D. Norton, *The Simplest Exorcism of Maxwell's Demon*, University of Pittsburgh, accessed April 22, 2025.

Conjectural thought experiments aim to identify a phenomenon that occurs during the experiment and then formulate a theory that accounts for it. The example Brown gives is Newton's bucket (Ibid., 38-43). In this thought experiment a bucket filled with water is suspended from a rope and set to spin. Initially

the water surface remains flat as the bucket begins rotating, but gradually the water climbs the sides of the bucket, forming a concave surface. When the bucket stops, the water continues its rotation before eventually returning to a flat state. Newton argued that this demonstrates the existence of absolute space, as the water's concave surface cannot be explained by relative motion between the water and the bucket.

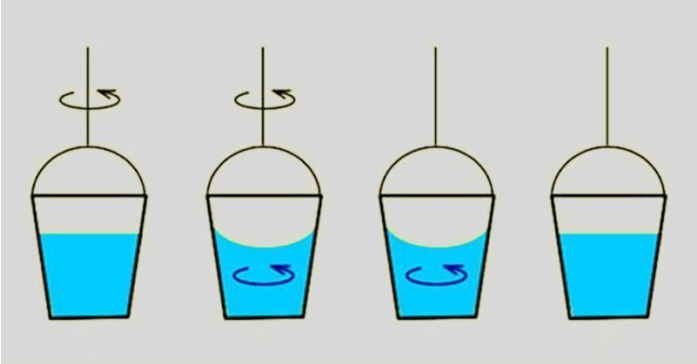


Figure 9: Illustration from Huzaifa Irshad, *Beyond the Spin: Newton's Bucket and the Riddle of Spacetime*, Journal of Young Physicists, April 26, 2024. Accessed April 22, 2025.

Direct thought experiments are similar to mediative thought experiments since they start with straightforward phenomena instead of an assumed phenomenon that occurs during the experiment; however, like conjectural thought experiments, direct thought experiments do not start with a clearly defined theory; they conclude with one. The example he gives of such a thought experiment is Galileo's one about falling bodies, which demonstrated that all bodies fall at the same speed, proving wrong the Aristotelian view that heavier objects fall faster. This exemplifies the destructive power of thought experiments: it exposed contradictions in the existing Aristotelian theory through purely conceptual reasoning, without requiring additional empirical observations to refute the established view (Ibid 39-41).

Platonic thought experiments are both constructive (direct) and destructive. Other than Galileo's falling bodies, he gives other two examples: EPR and Leibniz's account of *vis viva* (Cf. Ibid 41-43).

The EPR paradox is a thought experiment proposed by Einstein, Podolsky and Rosen that challenges quantum mechanics' completeness. In this scenario, two particles interact and separate, establishing quantum entanglement. Measuring one particle instantaneously determines properties of the distant particle, regardless of separation distance.

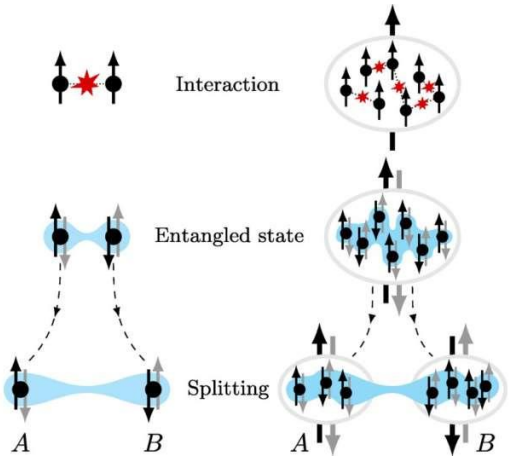


Figure 10: *Physical Review X* (2023). Accessed April 22, 2025.

According to Brown, such thought experiments allow us to grasp abstract entities and relationships in a Platonic realm of mathematical and physical laws. His theory is explicitly linked to Platonism, suggesting that laws of nature can be understood as relations among abstract entities. Just as we can apprehend mathematical truths through pure reason, Brown argues that specific thought experiments give us direct perceptual access to the laws of nature themselves (Brown 2004, 34).

This view offers a solution to the paradox of thought experiments. The Platonic answer is that we are not learning something new but instead directly perceiving abstract truths that were always there, waiting to be grasped by the "mind's eye".

Several philosophers have raised objections to Brown's view. Norton challenges the metaphysical assumptions of Brown's Platonism (Norton 1999, 129). Brown's Platonic account of thought experiments has faced criticism from both conceptual and empirical perspectives. While attempting to explain how thought experiments can produce new knowledge without new empirical data, Brown's appeal to a special Platonic faculty seems unnecessary. Norton argues that the alternative to his own view of thought experiments as arguments is "to suppose that thought experiments provide some new and even mysterious route to knowledge of the physical world" (Norton 1991, 129). He adds that he "can see no benefit in adopting a mysterious window view of thought experiments" (Norton 1991, 129).

Häggqvist points out an important problem in Brown's account: the small scope of his examples. Brown bases his epistemological claims on three cases: EPR, Galileo's falling bodies, and Leibniz's *vis viva* thought experiments. Moreover, "Of these three, the third is a calculation which reduces to absurdity Descartes' theory that conserved force is (something like) mass x

velocity and shows that it is quite generally *mv2* (which Leibniz called *vis viva*)." (Häggqvist 2009, 59).

Furthermore, according to Häggqvist, even these examples fail to support his Platonic theory. The "positive" or "constructive" aspects that Brown identifies in these thought experiments do not emerge from contemplating hypothetical scenarios. In the Leibniz case, the general conclusion about mv^2 stems from mathematical calculation rather than any unique insight gained through the thought experiment (Ibid). When Brown argues that these thought experiments support new theories, they often result in basically the negation of the theories they aim to challenge rather than providing substantial positive insights.

Häggqvist also argues that the other two exemplars of supposed Platonic thought experiments presented by Brown are most plausibly construed as modal counterexamples to Aristotle's theory on falling bodies and the Copenhagen interpretation (Ibid).

The main problem with Brown's account of thought experiments, though, as Norton points out, is the need to rely on Platonic metaphysics. If we can account for how thought experiments can be valuable without resorting to Platonism, we should keep a more naturalistic approach. As Brendel argues, "Intuitions are to be viewed as mental propositional attitudes accompanied by a feeling of certainty. Although spontaneous and non-inferential, they are not a priori in the sense of being irreducible or independent to empirically gained knowledge" (Brendel 2004, 102). So, if we want to argue that thought experiments have value for somehow prompting intuitions, we can choose to defend naturalistic accounts of intuitions.

3.1.2 The argument view

The argument view of thought experiments, mainly defended by John D. Norton, holds that thought experiments are "merely picturesque arguments" (Norton 1996, 334).

According to Norton, "Thought experiments are arguments which:

(i) posit hypothetical or counterfactual states of affairs, and

(ii) invoke particulars irrelevant to the generality of the conclusion" (Norton 1991, 129).

This view stems from an empiricist perspective that all knowledge must ultimately derive from experience. Norton notes:

> My account of thought experiments is based on the presumption that pure thought cannot conjure up knowledge, aside, perhaps, from logical truths. All pure thought can do is transform

what we already know. This is true with thought experiments: they can only transform existing knowledge (Ibid., 130).

Norton's central thesis has two key components:

1) The reconstruction thesis maintains that thought experiments can be understood as arguments rooted in either explicit or tacit assumptions. From this perspective, "belief in the outcome-conclusion of the thought experiment is justified only insofar as the reconstructed argument can justify the conclusion" (Norton 1996, 339).

2) The reliability thesis posits that thought experiments can only serve as reliable epistemic tools if they function as arguments – understood broadly – capable of supporting their conclusions. As Norton states, "if thought experiments can be used reliably epistemically, then they must be arguments (construed very broadly) that justify their outcomes or are reconstructible as such arguments" (Norton 2004b, 1143).

As an example, Norton reconstructs Einstein's rotating disk as an argument. This thought experiment explores the effects of special and general relativity on a rigid rotating disk. In this thought experiment, Einstein considered a large, rigid disk rotating around its central axis. He asked what happens to the disk's geometry as it rotates.

According to classical physics, nothing unusual would occur. However, Einstein showed that from the perspective of an observer on the disk, several strange phenomena would be observed: as the disk rotates faster, objects on its circumference would appear to contract in the direction of rotation due to special relativity's length contraction effect; the circumference of the disk would not follow classical Euclidean geometry. Here is how Norton reconstructed this thought experiment:

(D1) In Euclidean geometry, the measured circumference of a disk is p times its diameter.

(Premise)

(D2) The geometry of a nonrotating disk is Euclidean. (Premise)

(D3) The motion of a radial element on a rotating disk is perpendicular to its length so that

(according to special relativity) the length is unaltered. (Premise)

(D4) The motion of a circumferential element on a rotating disk is along its length so that

(according to special relativity) the length is contracted. (Premise)

(D5) Therefore, the circumference of a rotating disk is more than p times the measured diameter. (From D2, D3, D4)

(D6) Therefore, the geometry of a rotating disk is not Euclidean. (From D1, D5) (Norton 2004a, 50-51).

Norton's answer to the puzzle is that thought experiments do not generate new knowledge but rather make explicit what was already implicit in our existing knowledge through argumentation (Norton 2004a, 52). This maintains empiricist principles while explaining the thought experiment's apparent epistemic power. This also means that good thought experiments are good arguments, and bad thought experiments are bad arguments.

Others have supported this argument view proposed by Norton. Irvine, for instance, agrees that "thought experiments [...] are simply arguments concerning particular events or states of affairs of a hypothetical (and often counterfactual) nature which lead to conclusions about the nature *of the* world around us", as he argues that this is the only possibility to understand how thought experiments can "yield some new information about the world" (Irvine 1991, 150), thus responding to the puzzle.

However, the argument view faces important challenges. Michael Bishop argues that Norton's account cannot adequately explain cases where scientists disagree about a thought experiment's outcome. Bishop's central thesis is that the argument view "cannot make sense of the historical trajectory of certain thought experiments. In particular, it cannot account for episodes in which different thinkers disagree about a thought experiment's results" (Bishop 1999, 535).

Bishop identifies that the argument view is committed to the following thesis: "Two tokens of a thought experiment are tokens of the same thought experiment-type if and only if they are tokens of the same t-argument type" (Ibid., 539). Bishop argues that this thesis is false because "The t-arguments proposed by Bohr and Einstein were not type-identical, but the thought experiments they proposed were type-identical" (Ibid). So, even though Bohr and Einstein analysed the same thought experiment, they proposed two different arguments. Thus, if it is possible to construct different arguments based on the same thought experiments, this means thought experiments are not arguments.

Sorensen, too, argues that thought experiments are not arguments. He argues that thought experiments should be studied as if they were (real) experiments. Then he poses "a parity thesis": "thought experiments are arguments if and only if experiments are arguments" (Sorensen 1992,

214). Norton would now need to show that physical experiments are arguments to maintain his position about thought experiments. Treating thought experiments as arguments confuses their experimental features. Like physical experiments, thought experiments involve the manipulation of variables. Sorensen points out that thought experiments exhibit the same "tinkering" patterns as physical experiments. They often begin with an apparent counterexample to a theory, anticipate replies, and then present artificial variations designed to isolate specific factors.

This experimental methodology is distinct from the deductive or inductive structure of formal arguments. While thought experiments might suggest arguments or be accompanied, they function distinctly as experimental devices. They test hypotheses through idealised scenarios rather than logical deduction. Sorensen maintains that if a reduction of thought experiments to arguments were valid, similar reductions would be possible for "such intellectual instruments as calculation, interview, mnemonic aids, observation, paradigms, research programs, and similes" (Ibid., 214-15). This extensive reductionism suggests something fundamentally wrong with Norton's approach.

Brown is the primary opponent to Norton's view, arguing that specific thought experiments provide direct access to abstract laws in a way that cannot be reduced to argument, as seen earlier. He argues that even if thought experiments can be reconstructed as arguments, that does not mean they *are* arguments. Using an analogy about visual size judgments, he notes that "even a success rate of 100 per cent in reconstructing judgments of size does not refute the common-sense claim that we judge size by means of visual shape" (Brown 2004, 36).

Brendel, too, rejects that the reconstruction of thought experiments makes them arguments, arguing that even if they can be reconstructed as arguments, this does not identify them as such: "Although thought experiments can be reconstructed as arguments, I do not want to go as far as Norton and claim that all thought experiments really are explicit arguments" (Brendel 2004, 96). Furthermore, as analysed previously, Norton claims that "the actual conduct of a thought experiment consists of the execution of an argument" (Norton 2004a,50). However, Brendel argues that psychological evidence does not support this. Instead, it supports that "there are differences in the reasoning processes between thought experiments and arguments" (Brendel 2018, 290).

Finally, Brendel points out how Norton expands what counts as valid logical reasoning to sustain his argument view:

> Whereas Norton initially identified thought experiments with classical deductive and inductive forms of arguments, he later expanded the scope of logic to inference to the best

> explanation [...], the rules of Bayesian confirmation theory [...], and indeed to reasoning based on informal logic (Ibid., 286).

Norton even claims that thought experiments "are just governed by a 'very general notion of a logic'" (Norton 2004a, 54).

This definition makes Norton's thesis unfalsifiable: "Norton's Reconstruction Thesis is immune to falsification and thus amounts to a thesis that is simply vacuously true" (Brendel 2015, 287).

Michael T. Stuart notices this problem as well: If "argument" is stretched to include any cognitive process or heuristic that leads from premises to conclusions, then everything is an argument making the argument view trivially. He also identifies a direct contradiction between Norton's leading positions:

> There is a serious [therefore] internal tension between Norton's account of thought experiments according to which thought experiments are filled with irrelevant but picturesque details on the one hand, and his account of induction according to which the particular details are crucially important for justification on the other (Stuart 2016, 461).

Norton cannot consistently maintain that thought experiments contain mostly irrelevant details and that inductive inference critically depends on the specific details of cases.

Norton responds to these critiques by emphasising the broad nature of his conception of argument: "I do not maintain that thought experiments are arguments because of some belief in the self-discipline of thought experimenters to restrict themselves to a narrow canon of logic. Rather I have confidence that logic evolves to embrace any new, good argument forms that may emerge" (Norton 2004b, 1144).

The strength of the argument view lies in its ability to explain thought experiments' reliability and epistemic power without appealing to mysterious faculties. Its main challenge lies in fully accounting for the apparent novelty and transformative power that thought experiments sometimes show.

Before turning to middle-ground positions, it is worth noting that this chapter pays attention to accounts that preserve the epistemic significance of thought experiments without relying on strong metaphysical claims. While the extremes of Platonism and deflationism frame the debate, the focus here lies with those views that offer naturalistic explanations for how thought experiments

generate insight, especially those that will later help us assess whether films can do similar epistemic work.

3.2 Middle ground positions, or moderate inflationism

Having examined the opposing views of Brown's Platonism and Norton's argument, we now turn to more middle-ground positions that attempt to assess thought experiments' value without relying on either Platonic metaphysics or reducing them entirely to arguments. Davies provides a taxonomy distinguishing between "extreme" and "moderate" versions of both "deflationist" and "inflationist" approaches to thought experiments. I will leave deflationist approaches for later. Davies identifies Brown as an extreme inflationist, given that he claims that thought experiments provide a priori access to abstract Platonic laws. A moderate inflationist position provides an intermediate approach that recognises the distinctive cognitive value of thought experiments while preserving a naturalistic account. The moderate inflationist "claims that TEs allow the scientist to mobilise cognitive resources not available in the kinds of scientific reasoning celebrated by moderate deflationism" (Davies 2018, 520).

3.2.1 Modal reasoning and counterfactual thinking

Several philosophers have developed middle-ground positions on thought experiments that avoid both Brown's Platonism and Norton's reduction to arguments. Classical defences by Sorensen and Williamson offer accounts of how thought experiments can have epistemic value while remaining grounded in naturalistic frameworks. Sorensen and Williamson connect thought experiments to our modal and counterfactual reasoning capacity. They argue that thought experiments tap into our natural ability to reason about possibilities and necessities. Williamson, in particular, emphasises how counterfactual thinking in thought experiments parallels our everyday reasoning about what would happen in hypothetical situations. This suggests that thought experiments, rather than mysterious a priori insights or mere rhetorical devices, are applications of normal cognitive capabilities.

According to Sorensen, a thought experiment is "an experiment that purports to achieve its aim without the benefit of execution" (Sorensen 1992, 205). So, for him, thought experiments should be "studied as if they were experiments" (Ibid., 214). His main thesis is that "thought experiments evolved from ordinary experiments by a process of attenuation" (Ibid., 186). This

evolution occurred when experimenters noticed that experimental plans had epistemic benefits before execution (Ibid., 191).

Sorensen notes that thought experiments may sometimes be impossible to execute for several reasons, whether because they can be impossible due to technological or physical constraints or just because they are unethical (Ibid 197-222). So, although thought experiments happen in the mind and often cannot be executed, they are to be studied as real experiments as they have the same epistemic impact.

Sorenson also argues that thought experiments function as "necessity refuters" or "possibility refuters," testing the modal implications of theories through a structured logical process (Ibid., 135). Necessity refuters challenge theories that imply that particular possibilities are impossible. They establish that something ruled out as impossible by a theory is possible, thereby undermining the theory. He argues that necessity refuters consist of five key elements: the source statement (the theory being challenged), a modal extractor (drawing the relevant necessity claim from the source), a counterfactual (showing what would happen in a hypothetical scenario), an absurdity claim (establishing that the consequence is impossible), and a content possibility statement (affirming that the hypothetical scenario is indeed possible) (Ibid., 135-36). Gettier's counterexamples to the definition of knowledge "S knows that P" if "P is true", "S believes that P" and "S is justified in believing that P" (Gettier 1963, 121), or the justified true belief (JTB). These counterexamples exemplify necessity refuters. Gettier's thought experiment envisions a scenario where Smith has justified true belief about a proposition but lacks knowledge because his justification depends on a lucky coincidence. In this scenario, Gettier presents cases where a person has a justified true belief that still does not qualify as knowledge because the belief is true by mere coincidence rather than because of the justification. For example, if Smith justifiably believes Jones will get a job and has ten coins in his pocket, forming the belief "the man who will get the job has ten coins in his pocket" this belief may be true but for lucky reasons (Smith himself gets the job and happen to have ten coins). As Gettier states, JTB definition of knowledge "does not state a sufficient condition for someone's knowing a proposition." (Ibid., 123). By establishing the possibility of justified true belief without knowledge, Gettier effectively refutes the JTB definition.

Possibility refuters target theories that inappropriately imply that certain impossibilities are possible. These thought experiments demonstrate that something a theory claims is possible is impossible. Sorensen illustrates this with the "paradox of the stone," which challenges the concept of divine omnipotence: if God exists, then an omnipotent being exists who could make a stone too

heavy for himself to lift, yet this entails a contradiction (Sorensen 1992, 153-54). This thought experiment reveals a logical impossibility within the concept of omnipotence.

Sorensen's theory, thus, explains why thought experiments are useful even without empirical execution. While executed experiments get their power from new empirical data, thought experiments must work exclusively with existing information. However, this limitation does not make them epistemically irrelevant, just different. Thought experiments show that reorganising existing knowledge without gathering new data is also valuable (Ibid 213-14). Now, this account attempts to ground thought experiments in natural cognitive capacities rather than "mysterious" a priori ones. It rests on an evolutionary argument that our modal intuitions were selected for being reliable when dealing with practical problems of survival and navigation (Ibid., 253).

Timothy Williamson also offers an account of thought experiments that ground their epistemic value in our cognitive capacities. According to him, thought experiments are "valid arguments about counterfactual possibilities" (Williamson 2004, 207). While this seems like Norton's commitment to treating thought experiments as arguments, Williamson's approach is different since, rather than requiring reconstruction into explicit logical form, Williamson treats thought experiments as inherently argumentative structures that work through our natural capacity for counterfactual reasoning. This is also a naturalistic account that connects philosophical methodology with everyday reasoning, specifically with our ability to evaluate counterfactual conditionals. Williamson rejects the notion that thought experiments operate through an exclusive form of rational intuition. Instead, he demonstrates that they use the same cognitive capacities in assessing ordinary counterfactual statements such as "If the bush had not been there, the rock would have landed in the lake" (Ibid., 188).

Williamson analyses Gettier's counterexamples to the justified true belief theory of knowledge to show the logical structure of thought experiments. This structure involves establishing the possibility of a scenario, showing the consequences that would follow and deriving a modal conclusion from those hypothetical premises (Ibid 183-87). According to Williamson, thought experiments are reliable (or unreliable) in the same way as our cognitive processes for actual situations. Counterexamples, such as Gettier cases that occur in real life, function like hypothetical cases.

Sorensen and Williamson offer naturalistic accounts of thought experiments that avoid both Platonism and reduction to arguments, however their approaches differ. Sorensen's account is experimental, viewing thought experiments as attenuated versions of physical experiments that

have evolved to achieve epistemic aims without execution. His framework emphasises their role as modal refuters that test theories through structured logical processes. Williamson, in contrast, focuses more on counterfactual reasoning, arguing that thought experiments are exercises in evaluating counterfactual conditionals using the same cognitive capacities we employ in everyday reasoning.

Where Sorensen emphasises the structured, logical process of refutation, Williamson shows how the logical structure of thought experiments mirrors our assessment of ordinary counterfactual statements. Sorensen's evolutionary argument grounds the reliability of modal intuitions in natural selection, whereas Williamson more directly connects philosophical methodology to common cognitive processes. These differences reflect different approaches to naturalising thought experiments – Sorensen connects them to the experimental tradition, and Williamson connects them to everyday reasoning practices. Sorensen categorises thought experiments as necessity or possibility refuters, but we will see that thought experiments may serve additional purposes. Also, Williamson's account does not imply that all cognitive value of thought experiments should be reduced to modal intuitions or that all thought experiments must be formulated as counterfactuals; only that our ability to evaluate those hypothetical scenarios offers a cognitive explanation for their epistemic value.

Sorensen and Williamson root their accounts in naturalistic cognitive capacities. Sorensen's evolutionary framing leans toward a quasi-empirical justification of modal intuitions, and Williamson adopts a more formal account of counterfactual reasoning that ties directly into logical validity.

3.2.2 Mental modelling

The middle-ground positions developed by Sorensen and Williamson offer foundations for understanding thought experiments' epistemic value. However, other accounts move beyond viewing thought experiments primarily as modal refuters or counterfactual evaluations, instead focusing on how they engage our ability to construct, manipulate, and draw inferences from mental representations of situations.

This "mental modelling account" was primarily developed by Nancy Nersessian and Nenad Miščević. Their unique perspective offers a cognitive explanation for the functioning of thought experiments and their potential for generating new knowledge: "The general hypothesis is that

thought experimenting is a species of reasoning rooted in the ability to imagine, anticipate, visualize, and re-experience from memory" (Nersessian 2018, 310).

According to the mental modelling account, thought experiments construct and manipulate dynamic mental representations that preserve the relationships of the phenomena being reasoned about. Rather than proceeding through formal logical operations on propositions (as the argument view suggests) or providing special access to abstract truths (as the Platonic account claims), thought experiments engage our cognitive capacity for mental simulation and modelling abilities that evolved for understanding the physical world.

Nancy Nersessian regards thought experiments as "simulative model-based reasoning" rather than purely propositional reasoning (Nersessian 1992b, 291). She argues that thought experimenters use reasoning by manipulating mental models of the situation depicted in the narrative (Ibid 291-92). She explains that the main difference between thought experiments and propositional reasoning is that the former constructs and makes inferences from a mental simulation (Ibid 292). This simulation creates a dynamic mental model where the experimenter can explore the consequences of variables in ways that might not be possible in the real world.

Importantly, she argues that "thought-experimental narratives function in much the same way as other narratives" (Ibid., 295), and they have five main features:

1) They are presented to the public in a narrative form, sometimes as visual illustrations too, as they invite the reader to imagine a dynamic scene that includes time and causal sequences;

2) The sequence of events the reader is invited to follow works the same as in the real world, even with fantastical situations;

3) The thought experiment we are presented with is the refined version of it, not the failed versions – because, as she notes, they can fail. In Nersessian's account, failure occurs when mental models misrepresent constraints of the domain, incorporate flawed background assumptions, or abstract away details that are essential to understanding the phenomenon in question. For example, she points to Maxwell's demon thought experiment, which failed because it incorporated the incorrect assumption that molecules could be manipulated individually without expending work. This leads to the mistaken conclusion that the second law of thermodynamics could be violated (Ibid. 297-98).

4) Some physical characteristics of objects are not included when they have no purpose – for example, we could say that the colour of people's shirts in the Trolley Problem is immaterial to its result;

5) They are usually very compelling, so we feel no need to do them in the real world.

Nersessian resolves the puzzle of thought experiments by demonstrating how mental models incorporate real-world constraints: "The constructed situation inherits empirical force by being abstracted from both our experiences and activities in, and our knowledge, conceptualizations, and assumptions of, the world" (Ibid., 297). She further explains that thought experiments can generate new knowledge by creating novel models that integrate information from multiple sources.

While some philosophers have attempted to reduce thought experiments to arguments, Nersessian maintains that "the modelling function cannot be supplanted by an argument" (Nersessian 1992b, 297). She acknowledges that thought experiments can be reconstructed as arguments after they are done, however, she insists that "the argument is not evident until after the thought experiment has been constructed and executed" (Ibid., 297).

Miščević also argues that thought experiments involve constructing and manipulating mental models. He is also an empiricist, arguing that thought experiments are "processes of reorganizing or rearranging the old pieces of knowledge – old experiences, beliefs and the like" (Miščević 1992, 215). He draws upon Johnson-Laird's psychological theory of mental models. According to Johnson-Laird, "A model represents a state of affairs and accordingly its structure is not arbitrary like that of a propositional representation, but plays a direct representational or analogical role. Its structure mirrors the relevant aspects of the corresponding state of affairs in the world" (Johnson-Laird 1980, 98). Johnson-Laird established a distinction between propositional representations and mental models that Miščević applies to thought experimentation. As Johnson-Laird explains, "A propositional representation is a description of a state of affairs, which may be true or false. It is evaluated concerning a model representing that state of affairs" (Ibid., 108). Mental models, in contrast, represent information analogically where "structure is a crucial part of the representation" (Ibid).

A central feature of Johnson-Laird's theory that is essential to Miščević's account is that "a model represented in a dimensional space can be directly constructed, manipulated, or scanned, in any way that can be controlled by dimensional variables" (Ibid). This explains why thought experiments like Galileo's on free-falling bodies are so powerful – they allow the thinker to manipulate variables and "see" outcomes in ways that propositional representations alone cannot. As Miščević argues, "Many problems, particularly those involving spatial relations, are much easier to solve in mental models, than verbally" (Miščević 1992, 221).

Miščević characterises these mental models as cognitive structures that "purport to represent concrete situations, with determinate objects and relations" (Ibid., 220). These models possess a "quasi-spatial quality", distinguishing them from purely propositional or verbal reasoning.

The power of mental models in thought experiments derives from several characteristics. They are flexible, allowing for "ad hoc combinations and recombination of items". They integrate "chunked material, like schemas and scripts allowing for the use of stored knowledge organized in default hierarchies". Their non-propositional component is "particularly suitable for geometrical and visual geometry-based reasoning" (Ibid., 222). Thus, mental models "enrich the enquirers data basis by deploying pieces of knowledge which were not seen to be relevant before, and items of practical know-how which can be manifested only in concrete manipulation" (Ibid., 223). Because they are better at manipulating and mobilising concepts, they are more accessible than abstract propositional reasoning. Hence, they have heuristic value too.

Another important aspect of Miščević's account is his incorporation of default hierarchies from cognitive science. These mental structures allow reasoners to efficiently filter irrelevant factors, explaining "why the jump to conclusion is so unproblematic" (Ibid., 226). Like Nersessian, Miščević argues that when we engage in thought experiments, "the irrelevant (or seemingly irrelevant) factors have been neutralised and taken care of by our background knowledge, and its default hierarchies" (Ibid). This gives Miščević's account an advantage over Platonism, as it can account for successful and unsuccessful thought experiments. While Brown must regard failed thought experiments as "anomalies, random failures of sight on the part of the mind's eye" (Ibid), Miščević can attribute them to false theories embedded in our background knowledge.

By contextualising thought experiments within the broader human cognition, Miščević demonstrates how they can generate new insights without appealing to mysterious a priori access to abstract truths. As he concludes, his account "explains the nature of data, points to the skills employed in manipulating the items in the model, and accounts for heuristic fruitfulness and relative ease of the work in the laboratory of the mind" (Ibid).

The mental modelling account of thought experiments draws support from cognitive science research. This empirical foundation strengthens the account by grounding it in findings about human cognitive capabilities. As Nersessian notes,

> Modern advocates of mental modeling, too, speculate that the capacity is fundamental to the brain and that it evolved as a means for simulating possible ways of maneuvering within the physical environment [...]. Thus, facility with mental modeling is a combination of an

> individual's biology and learning and develops in interaction with an external, natural, social, and cultural reality, which gives credence to the idea that scientists could refine and extend this capability in service of creative reasoning about nature (Nersessian 2018, 312).

Research on discourse comprehension provides evidence that humans naturally construct mental models when processing narrative information. Johnson-Laird's work on mental models demonstrated that people represent described situations in ways that go beyond the propositional content of texts. These mental representations preserve spatial, temporal, and causal relationships between objects and events described in discourse.

3.2.3 Constructivism

Constructivism is also an empiricist account of thought experiments. Tamar Gendler can also be considered as defending a mental modelling account along her constructivism, arguing that through engagement with imaginary scenarios, thought experiments activate cognitive processes that may be unavailable in abstract theoretical reasoning.

Gendler's position can be characterised as a form of constructivism that accepts the cognitive role of thought experiments in reconfiguring conceptual understanding. This process cannot be fully equated with argumentation because it employs distinctive quasi-perceptual cognitive mechanisms. As she explains, "contemplation of an imaginary scenario may lead us to new beliefs about contingent features of the natural world that are produced not inferentially, but quasi-observationally; the presence of a mental image may play a crucial cognitive role in the formation of the belief in question" (Gendler 2004, 1154).

Gendler, as noted before, draws on empirical research in cognitive psychology. She rejects what she calls the "Dispensability Thesis" and the stronger "Derivativity Thesis." The Dispensability Thesis holds that "any good scientific thought experiment can be replaced, without loss of demonstrative force, by a non-thought-experimental argument" (Gendler 1998, 401). The Derivativity Thesis claims that "the justificatory force of any good scientific thought experiment can only be explained by the fact that it can be replaced, without loss of demonstrative force, by a non-thought-experimental argument" (Gendler, 1998, 401) and, as we can see, this can be identified as Norton's argument view.

Against these views, Gendler argues that the thought-experimental format plays "an indispensable role in the persuasiveness" of paradigm cases like Galileo's falling bodies thought experiment and that "a similar degree of persuasiveness could not be obtained on the grounds of

explicit argument alone" (Gendler, 1998, p. 398). This is because thought experiments engage cognitive mechanisms that abstract argumentation may not activate.

Through engaging with concrete imaginary scenarios, thought experiments can activate cognitive perspectives that might otherwise remain inaccessible through abstract theorising alone. This explains why thought experiments like Rawls's "Veil of Ignorance," Thomson's "Violinist", and various trolley problems have such philosophical power – they make specific cognitive perspectives available that reshape our conceptual understanding. Gendler's analysis of Galileo's falling bodies thought experiment demonstrates her approach. By examining this paradigm case, she shows that the standard argumentative reconstruction fails to capture what makes the thought experiment persuasive. The Aristotelian, committed to the view that heavier bodies fall faster than lighter ones, would not be convinced by a straightforward argument because there are multiple ways to avoid the apparent contradiction (Gendler 1998, 404-407). The thought experiment is powerful because it reconfigures the conceptual space by forcing the Aristotelian to confront tacit knowledge about physical reality that was not previously accessible through explicit reasoning.

Gendler characterises this process as an "*experiment-in-thought*" (Gendler 1998, 414), where the reader discovers their own tacit beliefs about how the world works. By contemplating the specific scenario – imagining the strapped-together bodies falling – participants activate tacit knowledge that was not previously available to them in theoretical form. She explains, "What this reveals is that the initial reconstruction of the Galilean thought experiment [...] fails to capture what is really doing the work in the case" (Ibid., 407).

Gendler demonstrates the distinctive epistemic value of thought experiments that a quasi-observational account explains. By highlighting the unique cognitive role of imagining particular scenarios, she explains why thought experiments are irreplaceable tools in science and philosophy, even when argumentative alternatives are available.

These moderate inflationist views offer a compelling framework for understanding thought experiments as cognitively rich but not metaphysically mysterious. For my argument, I will align most closely with constructivist and mental modelling approaches, particularly those developed by Gendler and Nersessian. While Gendler and Nersessian emphasise quasi-perceptual and simulative processes in traditional thought experiments, films provide actual sensory input that nonetheless generates similar mental simulations. Films are not quasi-perceptual like traditional thought

experiments; rather, they provoke the same kinds of simulative reasoning that Gendler and Nersessian identify as central to thought experiments.

3.3 Scepticism against thought experiments, or extreme and moderate deflationism

A straightforward solution to the puzzle of thought experiments is to adopt a sceptical position: thought experiments do not provide new or valuable knowledge.

David Davies characterises the most sceptical position toward thought experiments as "extreme deflationism", exemplified by Pierre Duhem (1954) and Carl Hempel (1965). According to this view, "scientific TEs are of at best heuristic value. A TE may suggest that physical reality has a certain feature, and may even provide the idea for a concrete experiment, but TEs cannot themselves teach us anything about the world" (Davies 2018, 519). According to Davies, for extreme deflationists thought experiments merely serve as useful heuristic devices that might generate hypotheses, but these hypotheses require independent empirical testing before they can yield knowledge.

Duhem famously claimed that "a thought experiment is no substitute for a real experiment" and should be "forbidden in science, including science education" (Duhem cited by Brown and Fehige 2023). According to this view, at most, thought experiments might generate hypotheses requiring empirical testing before yielding genuine knowledge.

However, recently, it has been suggested that interpreting Duhem as completely rejecting thought experiments requires revision. As Buzzoni demonstrates, this misreading has oversimplified the relationship between Duhem and Mach's positions (Buzzoni 2018). A close reading reveals that Duhem and Mach approached thought experiments from opposite starting points but converged toward similar conclusions. As Buzzoni notes, "Just as there was in Duhem's criticism of *experiences fictives* a tendency leading him to give voice to one of Mach's basic empiricist claims, so also there was in Mach's interpretation of thought experiments a tendency in the direction of Duhem's conventionalism" (Ibid., 31). Far from representing opposed positions, Duhem and Mach's views reveal complementary perspectives.

While Duhem's criticism that thought experiments cannot substitute for real experiments holds some validity in scientific contexts, this would be an overly restrictive view of thought experiments' epistemic value, especially in philosophy. As Buzzoni demonstrates, even Duhem's position was less radical than commonly thought. This critique also would fail to recognise the

importance that some thought experiments definitely had on the history of science. For example, Einstein's thought experiments on relativity, or Galileo's falling bodies preceded empirical confirmation and still proved to be indispensable for scientific advancement.

A less radical sceptical position is what Davies calls "moderate deflationism," exemplified by John Norton. This view accepts that thought experiments can tell us something about the world but denies that they have distinctive epistemic virtues. According to Davies, Norton's position and moderate deflationism, in general, contend that thought experiments are "epistemically unremarkable, merely colourful uses of our standard epistemic resources – ordinary experiences and the inferences to be drawn from them. TEs simply reorganise and make explicit what we already know about the physical world" (Davies 2018, 519). Williamson's approach might seem to fall into this same category, given that for him counterfactual reasoning is an ordinary cognitive capacity. For moderate deflationists, Davies notes, the narrative elements of thought experiments merely distract from the underlying structure that does the actual cognitive work. Davies characterises both extreme and moderate deflationism as fundamentally sceptical about the distinctive cognitive value attributed to thought experiments, though they differ in how completely they reject such claims.

3.3.1 Background conditions and lack of detail

Kathleen Wilkes presents one of the most systematic critiques of philosophical thought experiments. Her critique centres on using speculative scenarios to explore philosophical problems, particularly personal identity, arguing that such scenarios are "both problematic and positively misleading" (Wilkes 1988, 2). Unlike many philosophers who employ thought experiments as tools for conceptual analysis, Wilkes challenges their usefulness and validity, especially in the domain of personal identity, where bizarre scenarios like people splitting like amoebae or brain transplants are commonplace.

Wilkes's first objection concerns inadequate background conditions in philosophical thought experiments. Unlike scientific thought experiments operating within established frameworks, philosophical ones often "float free from any such constraints," presenting scenarios "without giving us any theoretical framework within which we can make sense of the fantasy" (Ibid., 8).

To illustrate this problem, Wilkes examines the Ring of Gyges example from Plato's Republic. In this thought experiment, a shepherd discovers a hidden chasm during a storm. He enters and finds a corpse wearing a gold ring. The shepherd takes the ring and leaves. Later, he twists the ring and becomes invisible. The shepherd experiments with the ring: turning the hoop inwards turns him invisible, outwards restores visibility. The shepherd exploits this power, becoming a royal messenger, committing adultery with the king's wife, assassinating the king and seizing the throne.

This thought experiment fails to specify crucial background conditions required to properly evaluate it, such as whether the ring's owner also becomes intangible. If not, "he might by mistake bump up against, and get arrested by, a policeman [...] Thus a potential criminal may yet have self-interested reasons for staying within the bounds of morality" (Ibid., 11). While Plato could ignore these details because his interest was moral rather than metaphysical, thought experiments about personal identity cannot similarly ignore their background conditions without becoming problematic.

Wilkes' second objection focuses on natural kinds and their relationship to conceptual analysis. She contends that:

> Whenever we are examining the ranges of concepts that do not pick on natural kinds, the problem of deciding what is or what is not 'relevant' to the success of the thought experiment is yet more problematic than the same question as it arises in science; and, unlike the scientific problem, it may not even have an answer in principle (Ibid., 15).

This issue is particularly severe for thought experiments about personal identity because persons are not natural kinds. Unlike terms that pick out natural kinds (such as "water" or "gold"), concepts like "person" lack the precision and theoretical background necessary to determine what variations are relevant or irrelevant in a thought experiment.

She further argues that scientific laws, especially those describing biological kinds like human beings, "are not disjoint and independent, detachable from one another [...]. They are interrelated, to varying degrees" (Ibid., 29). This interconnectedness means that changes to one aspect of reality cascade into other domains in ways that thought experiments typically fail to account for. When philosophers ask us to imagine scenarios like people splitting like amoebae or mind transplants, they isolate certain features from their necessary connections to other aspects of reality, rendering our intuitive responses unreliable.

In her critique, Wilkes specifically targets thought experiments concerning personal identity, such as brain transplants, hemisphere splits, people splitting like amoebae, mind transfers, and teleportation scenarios. She argues that these fantastical scenarios abstract away from the real world so drastically that they cannot inform our understanding of personal identity as it operates. She states in her preface, "Personal Identity has been the stamping-ground for bizarre, entertaining, confusing, and inconclusive thought experiments. To my mind, these alluring fictions have led discussion off on the wrong tracks" (Ibid., vii).

Instead of relying on such fantastical thought experiments, Wilkes advocates examining actual cases that challenge our concepts of personhood and personal identity, such as multiple personality disorders, commissurotomy (split-brain) patients, fugue states, severe amnesia, and various mental disorders. She contends that these real-world cases provide better insight into the nature of personal identity because they occur within a fully specified background of scientific understanding, avoiding the problems that plague fantastical thought experiments.

So, rather than employing speculative scenarios, Wilkes proposes focusing on "real people" and actual puzzle cases that "defy the imagination, but which we nonetheless have to accept as facts" (Ibid., vii). Her approach emphasises examining actual cases that challenge our conception of personhood, grounding philosophical inquiry in empirical evidence from psychology, neuroscience, and medicine, recognising that persons are fundamentally biological beings, and understanding that personal identity is embedded in a complex network of physical, psychological, and social facts.

Wilkes' critique challenges philosophers to ground their inquiries in empirical reality rather than relying on intuitions about fantastical scenarios that may tell us more about our conceptual limitations than the phenomena under investigation. By turning our attention to real cases and the complexity of human experiences, Wilkes suggests we can develop a more accurate understanding of personal identity than what emerges from thought experiments.

Wilkes contends that the more bizarre and unrealistic the thought experiment, the less utility it can have. She argues that "we cannot extract philosophically interesting conclusions from fantastical thought experiments" (Ibid., 1) because we face an unavoidable choice: either we picture them against the world as we know it (which deems them impossible by current scientific understanding), or we picture them against some entirely different background. If we choose the latter, "then we have the realm of fantasy, and fantasy is fine to read; but it does not allow for philosophical conclusions to be drawn, because in a world indeterminately different we do not know

what we would want to say about anything" (Ibid., 2). Her analysis of philosophical thought experiments on personal identity makes her argue that philosophical methodology should be grounded in the actual rather than the merely imaginable, especially when exploring concepts as fundamental to human experience as personal identity. As Wilkes succinctly puts it, "What actually can and does happen is usually more gripping than what the perfervid imagination dreams up from the philosopher's armchair" (Ibid., 1).

Johnathan Dancy critiques the use of thought experiments in moral reasoning too. Dancy asks "how one can reasonably come to a view about what it would be right or wrong to do in an imaginary case" (Dancy 1985, 144). As he observes, "an imaginary case has no nature beyond that specified in the description. Beyond what is explicitly offered, the imaginary case is indeterminate; there are no further 'facts of the matter'" (Ibid). Since imaginary cases are constructions with no reality beyond their description, any moral judgment about them risks being unfounded. The problem is not merely that our description is incomplete, but that there is no underlying reality to discover; there is simply nothing there beyond what has been explicitly specified. Thus, Dancy addresses the difficulties in moving from judgments about imaginary cases to judgments about actual ones. Even if we could somehow determine the moral properties of an imaginary case, applying this knowledge to an actual case remains problematic because the actual case contains complexities absent from the imaginary one. As Dancy explains, the actual case is "far richer in properties", containing more contextual details, emotional complications, and social factors than any imaginary case we might construct (Ibid., 148). This incompleteness undermines the applicability of imaginary cases to real moral dilemmas. Dancy questions the concept of "relevant similarity" that supposedly bridges imaginary and actual cases. The problem is circularity: "In order to determine whether A and B are similar in all morally relevant respects, or in all respects relevant to their moral properties, we would need first to decide on the moral properties of each" (Ibid., 147) Thus, we cannot determine whether imaginary and actual cases are similar without first knowing their moral properties, but the purpose of the comparison is to establish those properties themselves.

The problems constitute a challenge to the conventional use of thought experiments in ethics, especially within generalist frameworks that seek to derive universal principles from hypothetical scenarios. Despite these challenges, Dancy does not dismiss imaginary cases entirely. Instead, he reconceptualises their function within particularism.

For Dancy, moral principles should be understood as "reminders of the sort of importance that a property can have in suitable circumstances" rather than as strict rules (Ibid., 150). This creates space for imaginary cases as devices that reveal the potential moral significance of properties. In this reformulation, an imaginary case becomes "an abbreviated sketch of a case where a property can be seen to be important; where the importance it can have is revealed" (Ibid., 151). This allows us to understand why imaginary cases often make moral properties "easier to see" without requiring the problematic inferential move from imaginary to actual cases that generalist ethics demands.

For Dancy, moral principles should be understood as "reminders of the sort of importance that a property can have in suitable circumstances" rather than as strict rules (Ibid., 150). This creates space for imaginary cases as devices that reveal the potential moral significance of properties. In this reformulation, an imaginary case becomes "an abbreviated sketch of a case where a property can be seen to be important; where the importance it can have is revealed" (Ibid., 151). This allows us to understand why imaginary cases often make moral properties "easier to see" without requiring the problematic inferential move from imaginary to actual cases that generalist ethics demands. Dancy's theory thus is a middle ground in the debate over thought experiments in ethics. While rejecting the generalist's use of imaginary cases to establish moral principles that determine judgments in new cases, he preserves their value as tools that enhance moral perception. The imaginary case serves not to dictate a conclusion about an actual case but to illuminate possibilities of moral significance that might otherwise be overlooked.

James Wilson also identifies several challenges to achieving internal validity in thought experiments. Thought experiments function as short fictions and, thus, are underdetermined. The majority of questions about the world of the thought experiment remain unanswered, leaving different people to fill in contextual details in potentially divergent ways. This underdetermination creates problems for reproducibility (Wilson 2016, 136). For a thought experiments to demonstrate internal validity, different readers must arrive at similar judgements when presented with the same case, requiring not only that the case itself is clearly described, but also that readers interpret the morally relevant features of the case in similar ways to generate consistent responses.

Authors of thought experiments typically attempt to address underdetermination through conventions where the author stipulates all morally relevant features of the case. While this technique may enhance internal validity, Wilson suggests that it may undermine the value of thought experiments as good tools for ethics by constraining interpretations. As he states, "To

further spell out the implied convention, the author of the thought experiment has, by definition, specified all the elements of the case that are morally relevant. No morally relevant differences other than those that have been stipulated by the framer of the thought experiment apply to the situation" (Ibid., 137-38).

Wilson identifies a tension with this approach:

> Such techniques face a tension: in so far as the case is very sparsely described and is stipulated to have only the facts or features that its author says, then there is little to imagine, and it is difficult for its readers to experience the case as a genuine case. But the more difficult it is to experience the case as a genuine case, the less it is plausible to claim that thinking about the case has added something additional to thinking about the problem in purely abstract terms, and the less confident we should be that the judgements that arise from the case have clear implications about the actual world (Ibid., 138).

3.3.2 Physical scale and scientific constraints

Robert Klee argues that philosophers must better account for physical scale effects when constructing thought experiments, as neglecting these effects can undermine their conclusions. He argues that many philosophical thought experiments fail to consider how phenomena operate differently across vast scales of mass density, size, temperature and complexity in the physical universe.

Klee examines four major thought experiments to demonstrate his thesis: Putnam's Twin Earth and Brain in a Vat, Searle's Chinese Room, and Chalmers's Zombies Without Qualia. He argues that each of these overlooks crucial scale-dependent physical realities that could invalidate their premises. For example, the Brain in a Vat experiment ignores how actual brains swell and become dysfunctional when removed from skull pressure. In this thought experiment, Putnam describes a scenario where someone's brain has been removed and placed in a vat with nutrients that keep it alive. The nerve endings connect to a computer programmed to generate illusions of normal experiences. This challenges our ability to know whether our perceptions connect to an external reality. As Putnam argues, this scenario creates a sceptical problem about the possibility of knowledge about the external world.

The Chinese Room thought experiment challenges the claim that computers can understand language or possess minds simply by manipulating symbols according to a programme. Searle asks us to imagine himself inside a room filled with Chinese symbols and

equipped with an English rulebook to match these symbols. Despite not knowing any Chinese, Searle follows the instructions to respond to incoming Chinese messages with appropriate Chinese responses, passing the Turing test for understanding Chinese.[10] To outside observers who receive Chinese responses to their questions, it appears that the room (or Searle inside it) understands Chinese. However, Searle argues that despite the correct outputs, he has no understanding of the Chinese symbols he manipulates. He follows the rules without semantic understanding. Searle uses this thought experiment to distinguish between syntax (symbol manipulation) and semantics (meaning), arguing that computers operate only at the syntactic level. This defies the strong AI position that programmed computers have cognitive states similar to human minds (Searle 1980).

Chalmer's philosophical zombies, or p-zombies (Chalmers 1996) are hypothetical beings that are physically identical to human in every observable way but lacking conscious experience or phenomenal consciousness. This thought experiment serves as a counterargument against physicalism. P-zombies behave exactly like conscious human beings but would experience nothing internally or subjectively. The conceivability of such beings suggests that physical facts cannot fully explain consciousness, as two physically identical beings (a human and a p-zombie) would differ in the fact that the former has conscious experiences and the latter does not.

Now, according to Klee, the Chinese Room may fail to account for how linguistic competence requires specific neurological scale properties that cannot be replicated at room size (Klee 2008, 99-99). Thus, he argues for a "modal revolution" in philosophy, a shift away from logical possibility to consider physical possibility, aligning himself with Quinean naturalism and arguing that philosophy should be more informed by scientific understanding rather than a priori reasoning (Ibid., 101).

Klee's argument that philosophical thought experiments often neglect physical scale effects raises important concerns about naturalistic plausibility. However, this critique misrepresents what philosophical thought experiments aim to achieve. While Klee correctly distinguishes between logical and physical possibility, his recommendation to shift exclusively toward physical possibility overlooks the legitimate philosophical interest in exploring logical or conceptual possibilities that may not be physically realisable. Many thought experiments intentionally isolate concepts to test our intuitions about them independently of physical implementation. Klee's demand for scientific constraints overlooks that different thought experiments serve different purposes. When examining

[10] The Turing Test was proposed by Alan Turing in "Computing Machinery and Intelligence" (1950) to determine if a machine shows intelligent behaviour equivalent or indistinguishable from a human. In this test a human evaluator engages in natural conversations with two hidden participants, a human and a machine. If the evaluator cannot distinguish between the machine and the human, then the machine is said to pass the test.

conceptual questions, we can set aside physical limitations. For example, the Brain in a Vat scenario is not primarily concerned with neurophysiological details of maintaining an isolated brain, but with questions of epistemological scepticism, for example. The physical impossibility of certain scenarios does not necessarily invalidate their conceptual value. That said, Klee reminds that paying attention to physical constraints can strengthen thought experiments when they make claims about real-world application. This is important for thought experiments that explore phenomena like consciousness and personal identity. In these cases, it may be good to engage with relevant scientific findings.

3.3.3 Gap between hypothetical and real cases, and psychological limitations

Building on concerns raised in the previous section about thought experiments' connection to real-world contexts, Wilson provides an analysis of the "external" validity of thought experiments. He identifies two primary mechanisms by which thought experiments can lack external validity.

First, ethical judgements appropriate for thought experiments may depend on features not shared in other contexts. Here, Wilson challenges what he terms the "ubiquity thesis" that assumes that if a moral feature makes a difference in one context, it also makes the same difference in others. He counters this with the "principle of contextual interaction", which holds that moral properties can behave differently in different contexts – just like physical properties exhibit contextual behaviour (Ibid., 141). For instance, the moral significance of deception might vary. Deceiving someone to protect them from harm may carry different moral weight than deception for personal gain. A thought experiment presenting deception in isolation cannot capture how context fundamentally alters its moral character.

Second, ethical judgements appropriate in the thought experiments may presuppose causal structures that are different from those present in other contexts in relevant ways. Wilson illustrates this with examples from medical ethics and risk analysis, where thought experiments often assume degrees of certainty and causal simplicity that are not common in real ethical dilemmas (Ibid., 142).

While Wilson highlights the problems with the context for ethical judgements in thought experiments, Machery critiques the psychological foundations for thought experiments themselves. Where Wilson questions whether hypothetical scenarios can reliably map onto real world moral reasoning, Edouard Machery argues that the cognitive mechanisms underlying such judgement may be too weak when they are stretched beyond their evolutionary purposes.

Machery critiques philosophical thought experiments through a paradox: "Thought experiments are likely to be most useful for philosophical purposes when they describe situations that differ from the situations that elicit everyday judgments, but these situations are precisely those where the Parity Defense fails" (Ibid 2011, 205). The "Parity Defense" refers to the argument that because the same psychological capacities underlie everyday judgments and those elicited by thought experiments, the latter inherit the reliability of the former. The very features that make thought experiments philosophically interesting and helpful, as well as their ability to create novel scenarios that pull apart conceptually distinct but empirically correlated features, are the same features that undermine the reliability of our judgments about them. When philosophers construct cases that venture beyond the ordinary contexts of our psychological capacities, they cannot simply assume that our judgments about these cases inherit the reliability of our everyday judgments.

His "conditional" scepticism targets not all thought experiments, but those that describe situations far removed from the "proper domains" of our cognitive capacities – the everyday contexts in which our mechanism that form our judgements evolved to function (Ibid., 190-98). His main contention is that philosophical thought experiments routinely violate conditions essential for trustworthy intuitions.

First, they construct fanciful scenarios, like Gettier cases, zombies, trolley problems – that have little resemblance to real world experiences. As Machery notes, "[t]hought experiments in philosophy typically describe fanciful situations that are very remote from the situations that elicit everyday judgments, and thus we cannot rely on our memories of tried-and-true past judgments in everyday situations" (Ibid., 202). When philosophers ask us to imagine a justified true belief that fails to be knowledge or a physically identical zombie devoid of consciousness, they deliberately decouple features that ordinarily occur simultaneously in nature. This separation disrupts the heuristics we unconsciously rely on (Ibid., 204).

Second, thought experiments often include irrelevant narrative details that bias judgement. Unlike controlled empirical studies, which isolate variables, philosophical scenarios frequently include superfluous elements – names, settings, or backstories – that bias responses in ways unrelated to the theoretical question at hand (Ibid.).

Third, thought experiments exploit the "Ordinary Judgement Proposal", the idea that philosophical intuitions are produced by the same psychological capacities underlying everyday reasoning (Ibid., 193). While this proposal might legitimise thought experiments, Machery inverts its implications. If our cognitive capacities are adapted for real world contexts, their application to

contrived cases becomes epistemically problematic. The further a scenario strays from the proper domain of a cognitive capacity, the less confidence we should have in its outputs. This creates a dilemma: thought experiments are philosophically fruitful precisely when they violate the conditions that make ordinary judgments reliable.

Machery anticipates the "Expertise Reply" which claims that philosophers' training protects them from these pitfalls. However, he presents empirical evidence to undermine this defence. Machery argues that if philosophical expertise truly improves judgment reliability, we should expect this improvement to extend beyond thought experiments to philosophers' areas of specialisation. Specifically, if ethicists possess more reliable ethical judgments due to their expertise, and if these judgments have practical significance, we should expect ethicists to behave better than other philosophers. However, as he claims, Eric Schwitzgebel's research shows that ethical philosophers do not behave more ethically than other philosophers: they are 50% more likely to steal books from libraries, no more likely to follow basic norms of politeness, and no more likely to vote than other philosophers (Machery 2011, 209; citing Schwitzgebel's studies). Machery views this as indirect evidence that ethicists' judgments about ethical matters are not more reliable than those of other philosophers, which casts doubt on claims that philosophical expertise improves the reliability of judgments elicited by thought experiments. More importantly, there is no evidence that philosophers possess superior judgment about core thought experiment topics, even in ordinary cases, let alone in the exotic scenarios they devise (Ibid., 210). While Machery acknowledges this connection is indirect and that philosophers might fail to act on sound judgments for various reasons (weakness of will, inattention, etc.), he argues that expertise in argumentation does not translate to infallible intuitions.

The implications are significant. Machery does not suggest completely discarding thought experiments but emphasises the need for caution when interpreting their intuitive results as evidence. He advocates for complementing them with empirical data, and with theoretical strengths such as explanatory power, or situations that resemble everyday experiences (Ibid., 212). His scepticism remains circumscribed: it targets only those thought experiments that transgress the proper domains of our cognitive capacities, leaving ordinary judgments, and perhaps some philosophically modest hypothetical cases, intact (Ibid., 211). When thought experiments venture too far into the fanciful, philosophers should be wary of trusting the intuitions they provoke.

3.3.4 Logical limitations

Even when thought experiments remain within cognitive bounds, logical and methodological problems undermine their reliability. Sceptics like Peijnenburg and Atkinson contend that these experiments often fail to advance understanding; instead, they merely repackage the intuitions they claim to test, creating a vicious circle (Peijnenburg and Atkinson).

This circularity becomes especially clear when thought experiments are held to the standards of scientific evidence. Thagard, for instance, identifies five criteria that distinguish robust empirical methods from philosophical intuition pumps:

a) Reliability: scientific evidence produces consistent results across multiple applications. As Thagard notes, "A source of evidence is reliable if it tends to yield truths rather than falsehoods" (Thagard 2014, 294). While sources like systematic observations and controlled experiments are not infallible, they have consistently demonstrated a track record of producing truths.

b) Intersubjectivity: different observers can agree on the evidence. Scientific evidence does not depend on any individual's testimony but can be validated by different observers. In contrast, thought experiments show significant intersubjective variation, with philosophers often disagreeing vehemently about the implications of famous thought experiments (Ibid).

c) Repeatability: the evidence can be reproduced under similar conditions. Scientific evidence typically allows for the replication of results by the same or different persons at different times. While thought experiments can be repeated, there are "poor guarantees that similar results will occur at different times" (Ibid., 295).

d) Robustness: the evidence remains stable across different experimental setups. Thought experiments lack this robustness since they are "supposed to be evidence of a special kind that suffices on its own" (Ibid.).

e) Causal correlation with the world: the evidence has demonstrable causal connections to the phenomena under investigation. Whether from observation or experiment, scientific evidence maintains a causal connection to the world it describes. In telescopic observation, for instance, "reflected light enters the eyes of observers and stimulates their retinas, providing a causal influence from what is observed to the human observation" (Ibid). Philosophical intuitions, however, lack such causal connections to the world.

Unlike controlled experiments or observational data, thought experiments routinely fail on all counts. The same scenario, for example Jackson's Mary the colour scientist, can provoke

opposed conclusions, with some philosophers insisting Mary gains new knowledge upon seeing red and others denying it outright. Such persistent disagreements suggest the method's inherent flaws, not just the limits of individual cognition.

The Mary (the colour scientist), or the knowledge argument, made by Frank Jackson, challenges physicalist theories of mind questioning whether all knowledge can be reduced to physical information. Jackson presents Mary, a scientist who knows every physical fact about colour and vision while living in a black and white environment. The question follows: "What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she *learn* anything or not?" (Jackson 1982, 130). Jackson then argues that Mary would learn something new upon seeing colours for the first time, suggesting that there are facts about conscious experience (qualia) that cannot be deduced from physical facts alone. This would challenge the physicalist clam that all facts are physical, our conscious experience.

Now, the consequences of relying on thought experiments for getting evidence are articulated through what Thagard terms "the seven sins of thought experiments":

1) Generating falsehoods: thought experiments often lead to beliefs that contradict well established empirical findings. Thagard cites examples from philosophy of mind where thought experiments by Descartes, Kripke, Jackson, and Chalmers have supported mind-body dualism despite mounting neuroscientific evidence for mind-brain identity (Thagard 2014, 296).
2) Underspecifying situations: philosophical thought experiments typically provide insufficient detail to support reliable inferences. Thagard notes that they are "underspecified about what they are taking for granted," allowing philosophers to assume "that whatever they can imagine is actually possible" (Ibid., 297).
3) Justifying ignoring relevant evidence: thought experiments can serve as pretexts for dismissing empirical counterevidence. They justify ignoring empirical evidence that is, in fact, highly relevant, based on the assumption that philosophy should seek properties that hold in all possible worlds rather than just the actual one (Ibid).
4) Circular reasoning: thought experiments often bring in the conclusions they should establish. The method is "inherently circular, starting with a hypothesis and then making up pseudo-evidence to support the hypothesis" (Ibid., 298).

5) Blocking inquiry: by providing seemingly definitive answers, thought experiments can close off empirical investigation. When "used dogmatically," they imply "that alternative views are necessarily false," thereby impeding further investigations (Ibid).
6) Wasting time: thought experiments divert intellectual resources from more productive empirical approaches. Thagard laments seeing philosophers "wasting time on generating and responding to thought experiments rather than dealing with ethical, epistemological, and metaphysical questions using what is known about the world and the people in it" (Ibid).
7) Casting philosophy into disrepute: the questionable epistemic status of thought experiments can undermine philosophy's credibility within the broader intellectual community. This "final harmful effect" lends "credence to the dismissal of philosophy by scientists and others who are serious about understanding the world" (Ibid). Instead of philosophical methodology reliant on thought experiments, Thagard advocates for "natural philosophy" that aligns philosophical inquiry more closely with empirical science (Ibid., 299).

Thagard's critique presents a systematic challenge to the legitimacy of thought experiments. However, his objections rest on some problematic assumptions. First, Thagard applies scientific standards of evidence to philosophy without justifying doing so. The charges of circularity and lack of intersubjectivity overlook the productive role that disagreement plays in philosophical thought experiments. While initial responses to scenarios like the Trolley Problem or Mary may vary, these variations become starting points for philosophical dialogue rather than endpoints. The disagreements themselves reveal important features of our concepts. Also, claiming that thought experiments generate falsehood by contradicting empirical findings misconstrues their purpose, as they usually aim at exploring conceptual possibilities, connections and implications, not empirical facts, and thus they should be evaluated based on their relevance and philosophical fecundity rather than empirical accuracy. Finally, Thagard's criticism that thought experiments block inquiry and waste time by providing seemingly definite answers fails to acknowledge how they frequently open new investigations. The most productive thought experiments in philosophy generated decades of debate that advanced philosophical understanding. Gettier cases, for instance, did not just create controversy but improved our grasp of knowledge and justification, while the trolley problem has yielded valuable insights into moral psychology and the structure of ethical reasoning. So, while we should remain alert about the limitations Thagard identifies, his critique fails to

appreciate how thought experiments play an important role in philosophy, especially when instead of competing with, they complement empirical evidence.

The collective weight of these criticisms suggests that using thought experiments as evidence may distort rather than clarify philosophical discussions. While they can occasionally inspire new ideas, their tendency to deepen divisions, overlook empirical evidence, and promote circular reasoning calls for a more cautious approach.

My analysis of the major positions on thought experiments demonstrates that moderate inflationist approaches, particularly mental modelling and constructivist frameworks, offer the most philosophically interesting account of how thought experiments generate epistemic value. These frameworks successfully emphasise how thought experiments activate previous knowledge while avoiding both the metaphysical nature of Platonism and the reductive limitations of the argument view. Importantly, they preserve thought experiments' distinctive epistemic virtues without requiring problematic ontological commitments. Given this theoretical foundation, my defence of cinematic thought experiments will build on these insights to show how films can achieve the same epistemic work through their distinctive representational capacities.

Throughout my defence of cinematic thought experiments, I will also address some of the sceptical positions and arguments not yet addressed, since they can be viewed as what McClelland called "opportunities to achieve substantive insights into the distinctive epistemic virtues of the filmic philosophical thought experiment" (McClelland 2019, 112).

3.4 The role of intuition in thought experiments

The criticisms of thought experiments examined in the previous sections raise questions about the role of intuition in thought experiments. If thought experiments rely primarily on intuitive judgments, and these judgments are subject to the limitations and biases we have explored, what epistemic status should we attribute to intuition itself?

George Bealer defines philosophical intuitions as "intellectual seemings" that present themselves as necessary truths (Bealer 1998, 207). For Bealer, these intuitions function as a form of a priori evidence that cannot be reduced to empirical observation or conceptual analysis alone. Elke Brendel defines intuitions as "mental propositional attitudes which are accompanied by a strong feeling of certainty" (Brendel 2004, 96). These intuitions are not innate or independent of

empirical knowledge but are "shaped by our causal interactions with the environment" (Ibid., 102). This explains why intuitions can be relatively stable in familiar contexts, unstable or unreliable in novel or unfamiliar situations, and subject to modification through experience and learning.

Intuitions can function in several distinct ways within the specific context of thought experiments. The most obvious role of intuition in thought experiments is evaluating the described scenario. When considering trolley problems, for instance, our intuitive judgment that it would be permissible to divert the trolley but not to push someone onto the tracks provides the data that ethical theories must explain. Thought experiments often require judgments about possibility and necessity, too. Whether Mary, the colour scientist, learns something new upon seeing colour, we intuitively judge what is possible given specific initial conditions. Many thought experiments aim to clarify concepts by testing their application in unusual circumstances. Gettier cases, for instance, test our idea of knowledge by constructing scenarios where justified true belief seems intuitively insufficient for knowledge.

Among moderate inflationist positions, several philosophers have explored the role of intuitions in thought experiments without appealing to Brown's Platonism. These accounts seek to explain how thought experiments can generate new insights by drawing on our intuitive capacities while maintaining a naturalistic framework.

Brendel's approach represents a middle ground between rationalism and empiricism. Thought experiments are neither pure a priori reasoning nor simple empirical arguments. Brendel proposes four critical criteria to distinguish legitimate thought experiments from what Daniel Dennett calls "intuition pumps" (Ibid., 106).

First, researchers must carefully consider how changes in one aspect might affect other implicit assumptions when imagining a scenario. She critiques Putnam's *Twin Earth* thought experiment as an "illegitimate intuition pump" because it fails to account for how changing water's molecular structure would fundamentally alter the entire scenario (Ibid., 99).

Second, the specific case must justify the general conclusion: the particular scenario should not rely on superficial emotional appeals or misleading descriptions. Brendel suggests asking: "Does a structurally analogous example have the same intuitive plausibility and lead to the same consequences?" (Ibid., 99)

Third, counterfactual assumptions must be irrelevant. Fictional or idealised assumptions should not contradict the background theories being investigated. She illustrates this through

Einstein's thought experiments, showing how some counterfactual assumptions can be legitimately exploratory while others might invalidate the entire experiment (Ibid., 100-101).

Lastly, conceptual application in unfamiliar contexts requires caution. Like Machery's critique of thought experiments that venture beyond our cognitive proper domains, researchers must recognise that intuitions may break down when applying concepts to extreme or science fiction-like scenarios. This is particularly relevant in philosophical domains like personal identity, where thought experiments can reveal the limitations of our conceptual understanding (Ibid., 102-103).

Brendel emphasises that while thought experiments are indispensable for argumentation, they must be employed carefully to avoid the "dangerous suggestive power of misleading intuitions" (Ibid., 89).

Another important perspective comes from Ichikawa and Jarvis. Their central thesis is that thought experiment intuitions can provide a priori knowledge of necessary truths, contrary to Timothy Williamson's view that such intuitions are merely a posteriori counterfactual judgments. They begin by critiquing Williamson's counterfactual account, arguing that it fails to capture the modal status traditionally attributed to philosophical intuitions. As they note, "Williamson's counterfactuals [...] will vary in truth value between possible worlds" (Ichikawa and Jarvis 2009, 2), making such judgments inherently contingent rather than necessary. Instead, they propose an alternative framework based on truth in fiction, arguing that thought experiments work by allowing us to grasp necessarily true propositions through fictional narratives. Rather than making counterfactual judgments about what would happen in specific scenarios, they suggest we apprehend "thought-experiment intuitions through our grasp on truth in fiction" (Ibid., 1). This approach maintains the traditional status of philosophical intuitions as necessarily true and knowable a priori. A strength of their account is how it addresses naturalistic scepticism about intuitions. They argue that the same cognitive capacities that allow us to acquire everyday knowledge also enable us to gain a priori knowledge through thought experiments: "Whatever it is that explains our capacity for everyday knowledge should also be able to explain knowledge of thought-experiment intuitions" (Ibid., 14). By grounding thought experiment intuitions in our ordinary capacity to engage with fiction, they provide an account that does not conflict with naturalism, avoiding the need to posit mysterious rational faculties (Ibid., 1)

Daniel Dennett's analysis of "intuition pumps" provides a framework for understanding how intuitions function within thought experiments. Dennett initially coined the term to critique

John Searle's Chinese Room argument: "I coined the term 'intuition pump,' and its first use was derogatory. I applied it to John Searle's 'Chinese room,' which I said was not a proper argument but just an intuition pump" (Dennett 1995).

This carried a negative implication, indicating that thought experiments lacked rigour and were merely rhetorical tools. However, Dennett's thinking on this concept changed. He recognises that intuition pumps, while not logically conclusive arguments, serve an important function in philosophical discourse: "I went on to say that intuition pumps are fine if they are used correctly, but they can also be misused. They are not arguments; they are stories. Instead of concluding, they pump intuition. They get you to say 'Aha! Oh, I get it!'" (Dennett 1995).

Dennett observes that many of history's most influential philosophical ideas have been presented not as formally rigorous arguments but as vivid thought experiments:

> If you look at the history of philosophy, you see that all the great and influential stuff has been technically full of holes but utterly memorable and vivid. They are what I call 'intuition pumps' — lovely thought experiments like Plato's cave, and Descartes's evil demon, and Hobbes' vision of the state of nature and the social contract, and even Kant's idea of the categorical imperative (Dennett 1995).

In Dennett's view, these thought experiments are powerful because of their ability to capture the imagination and structure thinking: "I don't know of any philosopher who thinks any one of those is a logically sound argument for anything. But they're wonderful imagination grabbers, jungle gyms for the imagination. They structure the way you think about a problem. These are the real legacy of the history of philosophy" (Dennett 1995).

Dennett distinguishes between different types of thought experiments, noting that some are "analyzable as rigorous arguments": "Some thought experiments are analyzable as rigorous arguments, often of the form *reductio ad absurdum*, in which one takes one's opponents' premises and derives a formal contradiction (an absurd result), showing that they can't all be right." (Dennett 2013, 5). He illustrates this with Galileo's thought experiment about falling objects, reinforcing our earlier analysis but now emphasising how the logical form of *reductio ad absurdum* explains the thought experiment's epistemic power. This shows that rigorous thought experiments can advance scientific understanding without requiring non-naturalistic explanations. He also sees thought experiments as essential but in need of careful handling: "These thinking tools seldom establish a *fixed* point – a solid 'axiom' for all future inquiry – but rather introduce a worthy *candidate* for a

fixed point, a likely constraint on future inquiry, but itself subject to revision or jettisoning altogether if somebody can figure out why." (Ibid., 13)

Timothy Williamson argues that what philosophers call "intuitions" are not a special class of mental states, but just applications of our ordinary capacities for judgements (Williamson 2007, 216). As previously seen, he maintains that thought experiments engage our capacity for counterfactual reasoning which is the same one that we use in our everyday life. So, in this perspective, intuitions are not considered as special abilities, but nor discarded as unreliable either.

However, not all philosophers share this optimistic view of intuition's role in thought experiments, as we have already seen in sceptical positions regarding thought experiments. Among the sceptics, Thagard presents a naturalistic critique that challenges the reliability of thought experiments as sources of evidence. Rejecting Platonistic interpretations that characterise intuition as an exceptional faculty for apprehending abstract entities, Thagard proposes that intuitions are "neuropsychological reactions generated by unconscious processes operating in parallel" (Thagard 2014, 291). He hypothesises that intuitions emerge from three primary neural processes: encoding representation, neural binding, and interactive competition (Ibid). While potentially appropriate for domains of expertise where intuitions are formed through repeated interactions with regular phenomena, this neurological basis undermines the reliability of intuitions in philosophical contexts where such regular interactions are absent.

The majority of attacks against intuition in philosophy are made by experimental philosophy. Experimental philosophers have been demonstrating the unreliability of intuition in various cases, showing that intuitions vary between demographic groups, and that they are influenced by other factors that are obviously irrelevant to the truth of the content of the intuition (cf. Stich and Tobia 2018, 372). There have been studies conducted to show how philosophical intuitions vary across demographic groups, with language, with the order in which the thought experiments are presented, that they are subject to framing effects and by the physical and social environment in which the intuition is elicited (Ibid., 372-74).

In response to these attacks, the "expertise defence" has been made, which Machery explores, as we saw. This defence argues that while empirical studies may reveal variance or unreliability in non-philosophers' intuitions, these findings do not undermine traditional philosophical methodology because philosophers possess special expertise that makes their intuitions more reliable or accurate. So, it could be argued that philosophers' intuitions are less susceptible to irrelevant factors such as framing effects, order effects, or environmental conditions.

Philosophers' intuitions are more likely to contain accurate content than those of non-philosophers. Williamson addresses both the expertise defence and challenges from experimental philosophy. Rather than claiming philosophers possess superior intuitions, he rejects the special status often given to intuitions in philosophy, emphasising instead that what philosophers do when responding to thought experiments engages ordinary human cognitive capacities, as we have seen. This shifts away from mysterious "philosophical intuitions" toward ordinary human capacities for counterfactual reasoning.

Instead of claiming philosophers' intuitions are less susceptible to bias or more likely to track truth, Williamson's account focuses on skill development. He maintains that philosophical training cultivates general skills in handling abstract matters and evaluating counterfactuals (Williamson, 2011). He states "the expertise defence does not imply that a good philosophical education involves the cultivation of a mysterious sui generis faculty of rational intuition, or anything of the kind. Rather, it is supposed to improve far more mundane skills, such as careful attention to details in the description of the scenario and their potential relevance to the questions at issue" (Williamson 2011, 216).

This approach acknowledges experimental philosophy's findings while rejecting many of its conclusions. Where experimental philosophers interpret demographic variations in intuitions as undermining philosophical methodology, Williamson sees such variations as data to be explained by different approaches to counterfactual reasoning or different background theories. He argues that experimental philosophers have not demonstrated that philosophical cases are ineffective or that philosophers are incompetent in addressing them. As he points out, "the experimental critique presents evidence that philosophically untrained subjects perform poorly at thought experimentation" but fails to provide "significant evidence that thought experimentation is worse off in the relevant respects than the cognitive skills they acknowledge to be enhanced by training in philosophy" (Ibid., 226).

Instead of claiming philosophers have special access to truth, he emphasises continuity between philosophical and scientific methodologies. Both involve forming judgments about cases, considering counterfactual scenarios, and making inferences about the best explanation. Philosophical applications, he suggests, "inherit a moderate degree of reliability from the more general cognitive patterns they instantiate" (Williamson 2007, 5). This preserves a role for armchair methods without appealing to mysterious faculties or making unsupported claims about philosophers' superior intuitions. He reinforces this point by arguing that "to follow the

experimentalists' advice not to use such tests is to make philosophy less scientific, not more" (Williamson 2011, 229).

Despite these critiques, Dennett's position on "intuition pumps" suggests that thought experiments might retain value as conceptual clarification and reorientation devices. Thagard likewise preserves their "generative function" while rejecting their "evidential function." These moderate approaches point toward potential roles for thought experiments that acknowledge their limitations.

The role of intuition in thought experiments is still up for debate. This discussion shows both the strengths and weaknesses of intuition.

As critiques of intuition have accumulated, it has become clear that the epistemic power of thought experiments may not lie in intuition at all. This is important when considering artistic media such as film, where propositional reasoning is less central. If thought experiments are to be taken seriously in such contexts, we may need to reconceive or even replace intuition as their primary epistemic mechanism.

This is the direction suggested by thinkers like Catherine Elgin, who challenges the centrality of intuition and instead recognises the roles of understanding, exemplification, and cognitive engagement with complex representations. These alternative models open the door to a broader rethinking of thought experiments, not as intuition pumps, but as interpretive tools for reorganising conceptual frameworks.

Building on the moderate inflationist stance toward thought experiments, I will remain agnostic regarding the precise epistemic status of intuition. While acknowledging the various critiques of intuition from experimental philosophy and naturalistic perspectives, my account does not depend on resolving these debates. Instead, I focus on how cinematic thought experiments mobilise cognitive resources beyond intuitive judgement, including resources like mental modelling, exemplification, imagination and emotional responses.

3.5 Types of thought experiments

Having outlined the puzzle of thought experiments, the main positions regarding that, and the role of intuition in thought experiments, let us now examine the possible types they can be. While we have already encountered several of these distinctions in the discussion of various critiques and theoretical positions, such as destructive thought experiments (like Galileo's), modal

counterexamples (like Gettier cases), and constructive possibility-establishing cases, it is worth organising these types more explicitly. By examining thought experiments through their functional purposes, we can discern which types are most adequate for cinematic thought experimentation. Not all thought experiments have the same purposes, and their varying characteristics suggest that some may translate more effectively to film than others.

Thomas Wartenberg offers an interesting categorisation of thought experiments according to their functional purposes, highlighting how they can serve diverse philosophical objectives.

3.5.1 Counterexamples to philosophical theses

Thought experiments can challenge philosophical positions by revealing internal contradictions or limitations. These challenges take two forms, as identified by Sorensen: necessity refuters and possibility refuters.

Necessity refuters directly challenge theories by establishing that something that a theory claims is impossible is actually possible. A classic example is Gettier's thought experiments in epistemology (Gettier 1963). By presenting cases where an individual has a justified true belief but lacks knowledge due to lucky coincidence, Gettier refutes the necessity claim that "if knowledge is justified true belief, then necessarily, if a person has justified true belief that p, then he knows that p" (Sorensen 1992, 137).

Possibility refuters target theories that claim something is possible when it is impossible. The "paradox of the stone" is an example that challenges the concept of divine omnipotence. The thought experiment asks: Can an omnipotent being create a stone too heavy for itself to lift? This question reveals a contradiction in the concept of absolute omnipotence.

Wartenberg gives the example of the Knowledge Argument, or Mary the colour scientist, as providing a counterexample to physicalism. Another example of a thought experiment aimed at counterexemplifying physicalism is David Chalmers' philosophical zombies. This counterexample shows that the physicalist position fails to explain the explanatory gap between physical processes and subjective experience. For another counterexample, John Searle's Chinese Room thought experiment challenges computational theories of mind.

This category of thought experiments aligns with Brown's "destructive thought experiments." These thought experiments reveal contradictions in existing theories, similar to what Kuhn identifies as experiments that expose anomalies within established theoretical frameworks (Kuhn 1977). Brown presents Schrödinger's cat as a destructive thought experiment challenging

the Copenhagen interpretation of quantum mechanics. The thought experiment here is a counterexample that shows that the principles that work for subatomic particles lead to absurd consequences for everyday objects (Brown 1991, 24). Norton would analyse these as deductive arguments that expose logical inconsistencies in theories, demonstrating how the initial premises lead to contradictory conclusions.

3.5.2 Establishing possibility

Some thought experiments aim to conceptualise and establish the possibility of a scenario. An example would be Descartes' evil demon hypothesis (Wartenberg 2007, 60-61). Descartes' evil demon hypothesis establishes the possibility that a deceiver could manipulate our entire sensory experience. By methodically considering how each sense and even mathematical reasoning could be subject to this deception, Descartes demonstrates that radical scepticism cannot be ruled out.

For example, Hilary Putnam's Twin Earth thought experiment establishes the possibility that meaning might not be determined solely by internal mental states. In this scenario, there is a distant planet, a Twin Earth, that is identical to Earth with one difference: the substance called "water" on Twin Earth has a different chemical composition. Earth water consists of H_2O, Twin Earth's water consists of XYZ. XYZ maintains identical properties of H_2O at normal temperatures and pressures; it tastes like water, and the oceans are constituted by it. In this scenario, $Oscar_1$ on Earth and $Oscar_2$ on Twin Earth lack knowledge of water composition. They share identical beliefs regarding "water". Since people on both planets use "water" with the same mental associations but refer to different substances, this establishes the possibility that meaning "ain't in the head" (Putnam 1975), establishing semantic externalism in philosophy of language.

Similarly, Derek Parfit's teleportation thought experiment establishes the possibility that personal identity might be preserved through discontinuous physical existence. In Parfit's teleportation thought experiment, a person enters a cubicle and presses a button:

> Suppose that you enter a cubicle in which, when you press a button, a scanner records the states of all of the cells in your brain and body, destroying both while doing so. This information is then transmitted at the speed of light to some other planet, where a replicator produces a perfect organic copy of you. Since the brain of your Replica is exactly like yours, it will seem to remember living your life up to the moment when you pressed the button, its character will be just like yours, and it will be in every other way psychologically continuous with you (Parfit 2016, 94).

The replica brain possesses identical memories up until the moment of scanning; the character matches, and psychological continuity exists between the original and the copy. This scenario opens up the consideration that personal identity is maintained even in different bodies.

Williamson emphasises that such explorations are not mystical insights but applications of our ordinary capacity for counterfactual reasoning. These thought experiments function as "valid arguments about counterfactual possibilities" (Williamson 2007, 207), using our everyday ability to evaluate hypothetical scenarios.

When thought experiments aim to establish the possibility of a scenario, they engage with what Norton would consider arguments about modal claims. Though Norton views all thought experiments as arguments, these particular ones demonstrate that certain states of affairs are logically or conceptually possible.

3.5.3 Demonstrating impossibility

Conversely, specific thought experiments seek to illustrate the impossibility of a given concept. These are what Wartenberg calls failed thought experiments, which demonstrate an impossibility.

An example that he mentions is Quine's "gavagai". Quine's thought experiment depitcs a linguist encountering a previously unkown language, with only behaviour as evidence. When a rabbit appears and a native speaker says "gavagai", the linguist tries to translate it as "rabbit", but Quine points out various alternative interpretations that could fit like "animal", "white", "undetached rabbit parts" etc. This would show that a unique correct translation is impossible (Quine 1960).

From Norton's perspective, these thought experiments function as arguments demonstrating logical or conceptual impossibility. They represent the inverse case of possibility-establishing experiments but have similar logical structures. These could be considered cases of Brown's "destructive" thought experiments as they undermine claims by showing that certain concepts or scenarios cannot be coherently conceived.

3.5.4 Establishing necessary connections

Some thought experiments aim at establishing necessary connections between different concepts. One example is John Locke's thought experiment of a prince's mind being transferred to a cobbler's body to differentiate "person" from "man" (or human as animal, in a thesis about personal identity (Ibid., 62-63). Locke argues that the personal identity follows consciousness rather than the

physical body. The individual in the cobbler's body would be the same person as the prince because he maintains the prince's consciousness, memories and sense of self. This would establish the psychological continuity of personal identity.

Harry Frankfurt's thought experiment is a good example of this type of thought experiment. Frankfurt imagines a hypothetical controller who can intervene in decisions, ensuring that subjects cannot choose differently, and still never interferes with their original choice. The agent makes the choice freely, so no intervention occurs. Frankfurt concludes that responsibility stems from decision sources instead of possibilities: "The principle of alternate possibilities is false. A person may be morally responsible for what he has done even though he could not have done otherwise. The principle's plausibility is an illusion, which can be made to vanish by bringing the relevant moral phenomena into sharper focus" (Frankfurt 1969, 829-30).

Daniel Dennett's thought experiments about gradually replacing neurons with artificial components establishes necessary connections between functional organisation and consciousness. By examining how our intuitions shift through different replacement scenarios, Dennett establishes that consciousness must be necessarily connected to functional patterns rather than specific physical substrates, challenging dualist intuitions about the necessary connection between consciousness and non-physical properties (Dennett 1978).

These thought experiments align with aspects of what Brown calls "mediative" thought experiments, which help reach conclusions based on specific theories by illustrating conceptual relationships. Norton would reconstruct these as arguments demonstrating necessary conceptual connections, often as conditional statements showing that if certain concepts apply, others must necessarily follow.

3.5.5 Theory confirmation

The fifth type of thought experiment that Wartenberg analyses seeks to confirm a theory. Some thought experiments are designed to validate existing theories.

Arthur Danto's exploration of artistic identification through a thought experiment about mistaking a sculpture for a real object is a prime example (Wartenberg 2007, 64). Arthur Danto's thought experiments involve indistinguishable objects, one an everyday item and the other a work of art. For another example, Danto imagines two visually identical red canvases: one is merely a painted surface, while the other is titled something like *Kierkegaard's Mood*, inviting interpretation (Danto 1981, 1). This confirms his theory that art requires historical and theoretical context,

demonstrating how artistic identification operates within an art world of shared interpretive practices and theories.

Another example is Judith Jarvis Thomson's Violinist. In her thought experiment, a person wakes up connected to an unconscious violinist. The violinist suffers from a fatal kidney condition. The Society of Music Lovers determined that the person has the only compatible blood type. They kidnap the person and connect the violinist's circulatory system to them so that their kidneys filter toxins from both people. The Society acted without hospital authorisation, but disconnecting would cause the violinist's death. The connection must remain for nine months, after which the violinist will recover and separation can occur safely. Thomson asks whether there is any moral obligation in this arrangement for the person to continue to be connected to the violinist (Thomson 1971). The thought experiment confirms the hypothesis by showing how our intuitions about bodily rights remain strong even when the stakes involve another's survival, supporting a theory for bodily autonomy.

These thought experiments correspond to Brown's mediative thought experiments that illustrate or exemplify theoretical claims. Brown notes that they help "reach a conclusion based on a specific theory" (Brown 1991, 35-38). James Clerk Maxwell's demon is a thought experiment that challenges the second law of thermodynamics by imagining a scenario where entropy appears to decrease without expending energy. Maxwell proposed a tiny being who controls a door between two gas chambers, selectively allowing fast-moving molecules to pass in one direction and slow ones in the other, seemingly creating a temperature difference and reducing entropy. This raised the question of whether intelligent intervention could violate thermodynamic laws. However, the demon's information storage and erasure would itself produce entropy, preserving the second law. The demon must measure each molecule's speed, store this information, and periodically erase old data to make room for new measurements. This information processing, particularly the erasure step, requires energy and generates heat, thus increasing the system's total entropy even as the gas chambers become more ordered. The thought experiment thus illustrates the connection between information and thermodynamics by showing that information processing has thermodynamic costs.

Norton would view these as inductive arguments that support theoretical claims by showing how they make sense of particular cases. Kuhn might have considered these as thought experiments that help scientists work within established paradigms.

3.5.6 Theory construction

Some thought experiments actively aim at constructing entire philosophical theories. Plato's *Republic* and John Rawls's "original position" in *A Theory of Justice* (1971) are examples of such (Wartenberg 2007, 65).

Plato's *Republic* uses the extended thought experiment of constructing an ideal city to develop a theory of justice. By systematically imagining this city's education, social organisation, and governance, Plato constructs a complex theory of justice as harmony among parts, in the city and the soul. The thought experiment illustrates a pre-existing theory and actively develops one by examining hypothetical structures and their consequences.

John Rawls's "veil of ignorance" thought experiment constructs a theory of justice by imagining people choosing principles without knowing their future positions in society. Behind this veil, rational self-interested parties would choose principles that protect the least advantaged, leading to Rawls' difference principle. The thought experiment constructs a framework for understanding fairness as what would be chosen under conditions of impartiality, generating principles that were not evident before the experiment.

This category aligns with Brown's "constructive" thought experiments, particularly his "direct" and "conjectural" types that aim to establish new theoretical frameworks. These thought experiments, as Brown describes them, "aim to identify a phenomenon that occurs during the experiment and then formulate a theory that accounts for it" (Brown 1991, 38-43). They also correspond to what Kuhn describes as thought experiments that help scientists conceptualise nature in new ways. For example, Brown classifies Newton's Bucket as a conjectural thought experiment. Newton's spinning bucket of water established absolute space with its concave surface, even after the bucket had stopped moving. Norton reconstructs this as an argument about needing a reference frame independent of observable objects.

It is important to note that these categories are not mutually exclusive, and one thought experiment can be included in several categories simultaneously. Let us remember Brown's argument that Galileo's Falling Bodies is a "Platonic" thought experiment that destroyed Aristotle's theory and constructed a new one. When Galileo imagined dropping two objects of different weights connected by a string, he revealed a contradiction in Aristotelian physics: the combined system would both accelerate and decelerate relative to the heavier object alone. Beyond merely exposing this contradiction, Brown argues that Galileo's thought experiment also established a new principle:

that all bodies fall at the same rate regardless of weight. This dual function of destroying an old theory while simultaneously generating a new one makes it one of Brown's rare "Platonic" thought experiments that provide direct access to laws of nature (Brown 1991, 41). Without accepting the Platonic view of thought experiments, we can agree that some experiments may serve several different purposes.

This chapter explored the nature, value and limitations of thought experiments in science and philosophy. I began by situating thought experiments within a broader tradition, tracing their origins from pre-Socratic philosophy to contemporary science and philosophy.

We presented a debate between two primary perspectives on thought experiments. On one side, Brown's Platonic view argues that thought experiments can provide access to abstract truths, suggesting they offer a priori insights into the nature of reality. Conversely, Norton's argument view contends that thought experiments are essentially arguments that reorganise existing knowledge without generating new information.

Between these two extremes, I explored moderate approaches that ground thought experiments in human natural cognitive capacities. Sorensen and Williamson emphasise how thought experiments draw on our natural ability to reason about possibilities and counterfactuals. The mental modelling account, developed by Nersessian and Miščević provides a compelling explanation of thought experiments, arguing that they engage our capacity to construct and manipulate dynamic mental representations. The constructivist account defended by Gendler furthers this position by showing how thought experiments can activate cognitive processes that reconfigure conceptual understanding through quasi-perceptual engagement with imaginary scenarios.

The chapter also examines significant challenges to thought experiments. I pointed to problems of inadequate background conditions, scale effects and the gap between hypothetical scenarios and real world. I also examined the role of intuition in thought experiments. While some philosophers view intuitions as a special cognitive faculty, others like Williamson see them as ordinary reasoning.

The chapter also provides a taxonomy of thought experiments, drawing on Wartenberg's functional categorisation. These include counterexamples to philosophical theses, establishing possibility, demonstrating impossibility, establishing necessary connection, theory confirmation

and theory construction. This taxonomy reveals the diverse epistemic functions thought experiments can serve.

We suggest that thought experiments are cognitive tools that cannot be reduced to arguments or mystical insights. They represent a method of philosophical (and scientific) reasoning. By examining strengths and limitations of thought experiments, this chapter sets the stage for exploring how art, in general, and film particularly function as such.

Table 1: Positions on Thought Experiments in Contemporary Debate

Position	Extreme Deflationism	Moderate Deflationism	Moderate Inflationism	Extreme inflationism
Proponents	Duhem, Hempel	Norton	Sorensen, Williamson, Nersessian, Miščević, Gendler	Brown
Central thesis	Thought experiments serve only heuristic functions and cannot generate knowledge	Thought experiments are arguments in disguise with no distinctive epistemic features	Thought experiments engage distinctive but naturalistic cognitive processes (modal reasoning, mental modelling, constructivism)	Thought experiments provide *a priori* access to laws of nature through direct intuition

4. Artistic thought experiments and the "bold thesis"

Having surveyed the major philosophical approaches to thought experiments, we are now able to explore their artistic dimensions. In this chapter, I consider whether and how artistic works, particularly films, might function as philosophical thought experiments.

4.1 Artistic thought experiments

Having established a taxonomy of thought experiments, we can now examine how these categories might be applied to artistic media. While thought experiments in traditional philosophical contexts are typically designed to elicit specific intuitions about abstract problems, artistic thought experiments may do something unique. This raises important questions: Can films serve the same philosophical functions outlined in our taxonomy? Or do they need an expanded understanding of what philosophical thought experiments can be?

However, this focus on intuition obscures the potential that thought experiments might have. There can be alternative accounts that shift attention away from intuitive judgment. This chapter aims to rethink what thought experiments can be and how artistic media like film may fulfil this role. By moving beyond intuition, we open new possibilities for understanding how films can function as philosophical thought experiments. This approach allows us to see how films might counterexemplify philosophical theses, establish possibilities, demonstrate necessary connections, or even construct theories, functions identified in our taxonomy, while doing so through their distinctive audiovisual and narrative means. The key idea is that film's apparent limitations as a philosophical medium may actually constitute its unique philosophical strengths.

As Davies observes, thought experiments have distinctive epistemic value because they mobilise cognitive resources we already possess, "grounded in our experience of the world, which may not be available to us in any explicitly propositional form" (Davies 2019, 144).

Iris Vidmar offers an approach to understanding how artistic works can function as philosophical thought experiments. She articulates four distinct dimensions through which artistic works can engage in philosophy, offering a response to literary anti-cognitivists. As Vidmar explains, science fiction's "unbounded explorations of change" and "freedom from realistic conventions" create what theorists call "cognitive estrangement", a breach from familiar reality that parallels how philosophical thought experiments function (Vidmar 2012, 184). This distinguishes science fiction from realist literature, which adheres more closely to verisimilitude. Works in the realist tradition "cannot properly function as thought experiments in this first sense because they do not ask readers to engage with procedures or scenarios that 'have not been and cannot be enacted'" (Ibid., 183).

Now, according to Vidmar, there is a dimension that concerns how content is presented in artworks, specifically how they invite audiences to engage with counterfactual scenarios. Vidmar observes that science fiction bears the closest resemblance to traditional thought experiments. These works require an extra "intellectual step" from audiences, asking them to "imagine that p although you know that not p" (Ibid., 183).

Another dimension Vidmar identifies focuses on the cognitive aims of fictional scenarios. Certain works, particularly utopias and dystopias, are constructed with explicit philosophical purposes that parallel those of thought experiments (Ibid., 184-85). Drawing on Brown's taxonomy of thought experiments, Vidmar suggests that authors of works like Huxley's *Brave New World* or Orwell's *1984* intend their narratives to function similarly to philosophical thought experiments.

A third dimension Vidmar explores shows how literary works can function as thought experiments at the thematic level regardless of genre (Ibid., 186-87). She demonstrates that even realist fiction, though not similar to thought experiments in terms of structure, can mobilise conceptual knowledge and cognitive resources in ways that parallel them. In this dimension, science fiction and realistic literature tackle human issues, regardless of place or time. For example, she argues that Dreiser's *American Tragedy*, which is set against the backdrop of industrial developments, engages with "the question of agency, individual choices and free will, the reasons behind immoral behaviour and manipulation, the possibility of repentance, etc" (Ibid., 186).

One last dimension Vidmar introduces is between internal and external perspectives in literary engagement. This distinction allows audiences to approach fictional scenarios as observers (external perspective) or by placing themselves within the fictional world (internal perspective), fostering different modes of cognitive engagement (Ibid., 187-88). Vidmar notes that Davies also acknowledges this possibility when he writes

> Thompson's TE does differ from standard narratives in works of literary fiction in one respect. The reader is invited to imagine certain things occurring to her, rather than to a fictional character or group of fictional characters. But this is merely a rhetorical device on Thompson's part, and is in no way essential to the TE. We could replace references to 'you' with references to 'a woman', etc., without changing the way in which narrative functions (Davies 2007, 390).

So, while literary anti-cognitivists claim that literature can at best provide hypotheses rather than knowledge, as previously examined, Vidmar argues that hypotheses can have valuable cognitive benefits on their own (Vidmar, 189-200). Thus, she proposes that the cognitive value of hypotheses generated by literature lies in "their instrumental function, that is, in the impact they have for the way we account for the experience – cognitive and emotional – with the world" (Ibid 190).

This instrumental view of artistic hypotheses connects to broader questions about how fiction contributes to understanding. Building on this, we can also explore Catherine Elgin's idea of exemplification. While Vidmar specifically explores how literary works can function as thought experiments, Elgin questions how literature can contribute to understanding. Elgin rejects that understanding must be factive. While knowledge is widely accepted as factive, i.e., one cannot know something that is not true, Elgin argues that understanding operates differently: "I argue that a factive conception of understanding is unduly restrictive. It neither reflects our practices in ascribing understanding nor does justice to contemporary science, for science uses idealizations and models that do not mirror the facts. Strictly speaking, they are false" (Elgin 2007, 33). Now, if understanding does not need to be entirely factive, then works of fiction do not need to be dismissed as epistemically worthless simply because they describe scenarios that do not obtain.

Returning to the concept of exemplification (which we encountered earlier), Elgin applies Goodman's framework to explain how fiction can exemplify properties and patterns. According to Elgin, exemplification is a mode of reference by which an object highlights some of its features, bringing them to the fore (Ibid., 40). She explains that "[a]n item that at once refers to and instantiates a feature may be said to *exemplify* that feature" (Elgin 1993, 15). Just as a paint sample exemplifies a colour by instantiating and referring to it, a thought experiment or work of fiction can exemplify properties, patterns, or relationships: "Although fictions and thought experiments are not physical realizations of the phenomena they pertain to, like real experiments, they exemplify properties or patterns that they share with those phenomena. They thereby afford epistemic access to properties and patterns realized in fact" (Elgin 2014, 221).

This mechanism applies equally to scientific thought experiments and artistic works. As she argues, "Just as thought experiments are fictions in science, works of fiction are thought experiments in art. Both are vehicles for exploration and discovery, providing contexts in which features may be demarcated, their interplay examined, their implications drawn out" (Elgin 1993, 25). Through exemplification, artistic works can highlight patterns, relationships, and properties that might otherwise remain obscure, enabling audiences to recognise and attend to these features. Following Goodman, Elgin argues that cognitive progress often consists in reconfiguration: "in reorganizing a domain so that hitherto overlooked or underemphasized features, patterns, opportunities, and resources come to light" (Elgin 2002). Literature excels at such reconfiguration, enabling us to see familiar phenomena in unfamiliar ways. Elgin argues that certain non-literal representations, which she terms "felicitous falsehoods", can enhance our understanding precisely because they strategically deviate from strict accuracy, allowing us to grasp important patterns and principles that might otherwise remain obscure (Elgin 2004, 114-18). According to her, science frequently employs these felicitous falsehoods: "Science routinely transgresses the boundary between truth and falsehood. It smoothes curves and ignores outliers. It develops and deploys simplified models that diverge, sometimes considerably, from the phenomena they purport to represent" (Ibid., 113).

Here, Elgin's concept of exemplification helps explain literature's epistemic value. An epistemically effective fiction "exemplifies certain features that obtain in reality but are, or may typically be, difficult to discern in their natural setting" (Elgin 2019, 5). Once identified in fiction, we gain "epistemic access to them, and can recognise them and appreciate their importance when we encounter them elsewhere. Elgin views thought experiments along a continuum that ranges from austere scientific scenarios to rich literary works: "Although stereotypical thought experiments are austere, there is a continuum of cases from Maxwell's demon and trolley problems through the myth of the cave and Emile to 'didactic fictions' like *Animal Farm* and *Uncle Tom's Cabin*, to *Middlemarch* and *Oedipus Rex*" (Elgin 2014, 240). Thus, rather than making restrictive boundaries between thought experiments and literary fiction, Elgin suggests that they operate on the same principles, with differences being a matter of degree rather than kind. All these can advance understanding by exemplifying features they share with reality, making those features salient and accessible.

It is precisely by departing from literal truth that thought experiments and fiction can advance understanding: An experiment deliberately departs from nature in order to advance an understanding of nature. Rather than invalidating the experiment, this departure is what enables it to disclose barely detectable, or standardly overshadowed aspects of nature" (Ibid., 226). This helps explain why some idealisations are preferable to the truths they approximate. For instance, in fluid dynamics, scientists prefer a solvable first-order approximation to an unsolvable second-order equation that more accurately represents reality (Elgin 2004, 122). The approximation is more fruitful precisely because it sacrifices some accuracy for tractability. Similarly, fiction writers can isolate patterns by limiting the scope of their focus: "Because the relations among the members of three or four suitably characterized families are sufficiently complicated, and the demands of village life sufficiently mundane, [Austen's] stories can exemplify something worth noting about ordinary life and the development of moral personality" (Elgin 2014, 233). So, rather than measuring artistic works against the standards of explicit argumentation, we can appreciate how their aesthetic and narrative properties enable them to contribute to understanding through exemplification, making salient aspects of reality that might otherwise remain unnoticed or underappreciated.

While Elgin focuses on exemplification as an epistemic mechanism of fiction, other philosophers examine how readers engage with fictional content. Eileen John argues that works of fiction can serve as legitimate sources of conceptual knowledge by prompting readers to examine and refine their understanding of concepts. Rather than simply presenting readers with new concepts to adopt, fiction generates conceptual inquiry by raising "questions she sees as likely to occur within 'ordinary' responses to the story" (John 1998, 337). She challenges critics like Lamarque and Olsen, who claim that literary interpretation operates in an "alternative realm of application" separate from real-world conceptual understanding (Ibid., 342).

Instead, John contends that our conceptual engagement with fiction is continuous and can meaningfully inform our broader conceptual competence.

John develops this argument through several important claims. She claims that fiction can raise legitimate conceptual questions through readers' natural interpretive responses rather than requiring specialised philosophical analysis. Using Grace Paley's story "Wants" as a case study, John demonstrates how a straightforward reading raises complex questions about the nature of desire and wanting: "What kind of wants does this woman have? What kind of criticism is it when the ex-husband says she will always want nothing?" (Ibid., 337). She also affirms that the literary qualities of fiction actively support, rather than distract from, conceptual thinking. John argues that features like narrative context, character development, and symbolic imagery help readers explore conceptual questions in uniquely productive ways. The story's literary elements, from the ex-husband's bitter criticism to the symbolic sycamore trees, create an environment for examining "how a certain kind of desire, the kind this narrator may have, would work" (Ibid., 344). Finally, she states that fiction may sometimes employ genre-specific conceptual frameworks, but readers can generally distinguish between genre conventions and genuine conceptual thinking. John acknowledges that some literary judgment may be "tailored to meet the conventions of certain kinds of novels" (Ibid., 341) but maintains that readers can recognise when they are exercising "standard conceptual competence in a thoughtful, responsible way" (Ibid., 342).

John suggests that dismissing fiction's philosophical value based on its literary qualities misconstrues how conceptual understanding actually develops through engaged reading experiences. As John concludes, investigating "how that philosophical dimension is integrated into literary experience can enrich our appreciation of literature, without narrowing it into a tool of philosophy" (Ibid., 346).

John's emphasis on readers' conceptual engagement raises questions about the different modes of imaginative thinking involved in artistic thought experiments. Elisabeth Camp also argues that artistic works can function as philosophical thought experiments by distinguishing between two forms of imagination: pretence and metaphorical seeing-as. Camp demonstrates that different kinds of imaginative engagement serve distinct epistemic functions, each with unique potential for philosophy.

"Pretence" is the imagination that manipulates content by "altering what is so, in order to produce a new world in the imagination" (Camp 2009, 116). This corresponds to the traditional conception of thought experiments as counterfactual scenarios that test concepts. "Metaphorical seeing-as" is imagination that manipulates the mode of presentation of content by altering how we structure our thoughts. To illustrate this distinction, Camp analyses what happens when we interpret "I am Anna Karenina" as a pretence or metaphor. In pretence, we transform ourselves imaginatively to become like Anna by endowing ourselves with properties Anna possesses. In metaphorical interpretation, we use Anna to structure our understanding of ourselves as we are, highlighting Anna-like character features while downplaying others (Ibid., 112).

Camp identifies three differences between these imaginative activities. First, the "direction of fit". In pretence, real objects serve as props for imagining something else; in metaphor, the imagined frame serves to restructure understanding or a real object (Ibid., 113). Second, the "mechanism" of fit. Pretence endows subjects with properties possessed by the framing object, while metaphor highlights properties of the subject that are merely similar to the framing object's properties (Ibid). Finally, the phenomenological experience. Pretence requires complete immersion, and metaphor involves distinctive "twofoldness" where we simultaneously maintain awareness of both the subject and frame as distinct entities (Ibid.)

Camp's analysis reveals how artistic works can function as thought experiments more complexly than previously recognised. Rather than limiting artistic thought experiments to pretence (counterfactual scenarios), she demonstrates how artworks often employ both pretence and metaphorical seeing-as simultaneously. Camp concludes that "pretense and metaphor are no substitute for rational argument, but they can provide grist for our argumentative mills, and spotlights into domains where we are otherwise groping in the dark" (Ibid., 128).

Camp's distinction between pretence and metaphorical seeing-as provides a foundation for examining how these imaginative processes operate specifically in film. David Davies also tries to establish what kinds of relationships there are between thought experiments and art in general, and more specifically, film – a focus to be delved into later. As previously examined, Davies differentiates responses to the epistemological puzzle of thought experiments by organising them along a spectrum, from extreme deflationism to extreme inflationism. Davies establishes his own moderate inflationist perspective, which aligns with constructivism and mental modelling accounts (Davies 2018, 520).

Davies' moderate inflationist position recognises that thought experiments have epistemic value precisely because they draw upon "cognitive resources we already possess, grounded in our experience of the world, which may not be available to us in any explicitly propositional form" (Davies 2008, 17). This is particularly applicable to the analysis of artistic works as philosophical instruments. Davies contends that the narrative elements in thought experiments are not merely decorative but essential to their function, as they activate tacit, experientially-grounded knowledge that shapes understanding. Thus, the cognitive value derives not from explicit arguments but from the mobilisation of tacit knowledge during engagement with the narrative.

When applied to artistic works generally, Davies' moderate inflationist view indicates that "mental modelling" occurs during engagement with fiction. Davies writes that "the mental models through which readers comprehend fictional narratives also provide, through their mobilisation of tacit or unarticulated knowledge of the world, a means of testing those claims to knowledge of the actual world that theorists have located in fictional narratives" (Davies 2007, 44). This account of mental modelling aligns with the positions of Nersessian and Miščević, who emphasise that thought experiments function by enabling the construction and manipulation of mental models.

Based on Richard Moran (1994) Smith distinguishes between two forms of imagining in our engagement with artistic works (2006), a distinction that Davies uses. First, there is what he terms "hypothetical imagining," which involves considering what would be the case if certain conditions were obtained. This form of imagining is central to traditional philosophical thought experiments. Second, "dramatic imagining" involves a more immersive engagement with fictional scenarios, including emotional and perceptual aspects of experience (Davies 2019, 151). Davies suggests that dramatic imagining, characteristic of our engagement with artistic works, enhances rather than detracts from the philosophical potential of these works.

This difference is important for grasping how art can serve as a way to explore ideas. Unlike standard philosophical thought experiments, which often rely on imagining scenarios, art invites us to engage through dramatic experiences. Davies suggests that this kind of experience can offer insights that we cannot achieve through hypothetical thinking alone.

Davies' account offers a helpful way to understand how various artistic works can act as thought experiments. His moderate inflationist view helps us understand how these works can advance our philosophical understanding without just being used for logical arguments. This perspective values the rich details and emotional experiences found in art as important parts of philosophical exploration, as I will explore further.

In sum, artistic thought experiments mobilise cognitive resources through mental models that activate tacit knowledge unavailable in propositional form. For instance, understanding why a particular moral dilemma feels compelling or recognising how a character's emotional state shifts cannot be fully captured in abstract propositions alone. Their philosophical power derives from exemplification, highlighting patterns by both instantiating and referring to them, affording access to otherwise obscure features of reality. As "felicitous falsehoods," they enhance understanding by strategically deviating from literal accuracy. Operating across multiple dimensions simultaneously, artistic works engage with counterfactual scenarios and philosophical concepts regardless of their narrative structure. They employ both "pretence" and "metaphorical seeing-as," allowing us to test concepts.

Traditional and artistic thought experiments exist on a continuum rather than as distinct categories. Both advance understanding by making features salient, though through different means. There is good reason to think that artistic works may be able to perform various philosophical functions identified in our taxonomy, from counterexemplifying theses to constructing theories through their distinctive aesthetic qualities. What traditional philosophy might view as limitations constitute strengths, enabling artistic works to reconfigure understanding in ways abstract discourse cannot, something to be explored in detail for the case of film.

4.2 The rationality objection and the "bold thesis"

4.2.1 The rationality objection

Davies' provides a strong foundation for understanding how artistic works, including films, might function as philosophical thought experiments. However, this alone does not fully address a fundamental challenge to the idea of film as philosophy: even if films can present thought experiments, why should we use films for this purpose when traditional philosophical methods are available? So here we need to address the following questions: Is it possible for a film to present a thought experiment that holds the same or an analogous value as a written one? Does it effectively communicate concepts with the same clarity and certainty? Are there specific aspects of a visual thought experiment that set it apart from a written one and make it particularly compelling? Or as Tom McClelland put it: "Granting that some films can act as philosophical thought experiments, how (if at all) can they make a valuable contribution to philosophical enquiry?" (McClelland 2019, 95).

There is an easy way to solve this problem, and the generality and the explicitness objections: to maintain that films can present explicit arguments. Films can do so in various ways; for example, a film that is just a record of a philosophy seminar/lecture/presentation where philosophers are making explicit arguments. Wartenberg imagines a film called *Justifying Difference*, just a record of a lecture given by John Rawls (Wartenberg 2006, 128). Another real example he gives is *Waking Life* (Richard Linklater, 2001), where Robert Solomon, as an animated "character", gives a lecture and talks about existentialism. Other examples could be *The Pervert's Guide to Ideology* (Sophie Fiennes and Slavoj Žižek, 2012) or the American TV series *The Good Place* (2016-). However, as Livingston pointedly notes, this is not what philosophers want when arguing for film as philosophy. They want a film that philosophises through cinematic techniques. These recorded lectures are philosophy *filmed*, not film *as philosophy*. The cinematic medium is contingent on the philosophical content – the arguments could be equally presented in audio or text form. Unlike documentaries, which employ editing, framing, and visual rhetoric to construct meaning, filmed lectures treat the cinematic apparatus merely as a recording device. For film to philosophise, the medium itself must participate in the philosophical investigation.

This challenge becomes particularly pressing when we consider Livingston's rationality objection, which questions why we would use film for philosophical purposes when more traditional means are available. Before examining how films might function as philosophical thought experiments, we must address this challenge to film's philosophical potential and understand what makes cinematic techniques philosophically significant.

The relationship between film and philosophy faces several significant problems, as discussed in previous chapters. Among these, the rationality objection poses a particular challenge. The rationality

objection is as follows: "If we in fact believe a more efficient means to our goal is available, would it not indeed be irrational to pass it by?" (Livingston 2006, 17). That is, if there are more effective means – in this case, traditional means of doing philosophy, such as texts or conferences – to more fully achieve our epistemic goals, why resort to a supposed less effective means, such as film? Using the most effective means for the objectives in question would be more rational.

This objection is important when we assess the evolution of philosophical methods. The classical tools of philosophy, such as written reasoning, formal logic, and structured essays, have developed over centuries to enhance clarity, precision, and rigour. The rationality objection argues that opting for a less refined tool over well-established methods is like intentionally selecting an ineffective way to reach our goals, assuming philosophy inherently possesses epistemic aims.

To further understand this problem, Livingston suggests an analogy with the use of a screwdriver to tighten a screw *versus* the use of a coin:

> If you know you can quickly, efficiently, and very effectively tighten a screw with a screwdriver that is ready to hand or laboriously and imperfectly tighten it with a coin, it would not be irrational to prefer the coin (assuming all other conditions are equal)? (Livingston 2009, 56).

This analogy emphasises two important points regarding the objection. First, it suggests that traditional philosophical methods – such as written texts and formal arguments – are more precise and efficient tools for conducting philosophy. Second, it implies that the limitations of film as a philosophical medium make it a less practical choice.

This analogy appears to appeal to common sense: why would someone choose a less effective tool when a better one exists? However, this view raises important questions. First, how do we judge the effectiveness of a philosophical tool? We might need to rethink the idea that written texts are always the best choice for philosophical work.

Livingston offers a reply to this objection. The two media (philosophical texts and film) do not have to be mutually exclusive; we are not forced to choose between one medium or another. We could further argue that posing such a dichotomy is a false dilemma. Using the screwdriver analogy, we would not have to choose between using a screwdriver or a coin: we could use both instruments to tighten a screw. Thus, "the idea that we must choose between doing philosophy with film and doing philosophy with linguistic and conceptual tools [...] is not accurate, since it is not what is at issue" (Livingston 2009, 56). The idea of complementary tools suggests that different methods can enhance our understanding of philosophy. For instance, films might be great for illustrating thought experiments, while traditional texts are better for logical analysis.

However, an important note from Livingston is that one could fulfil the purpose of tightening the screw only with the use of the screwdriver; that is, one could fulfil the objective of philosophising using only

the most traditional means; yet this is not incompatible with the idea that the use of a coin can help in the task, that is, a film can assist in the task of philosophising: "Any of the choices would be rational." So, one might think it is a rational strategy to use film as a helper in the "overarching project of philosophical pedagogy and research" (Livingston 2009, 57).

This response highlights an important point about how we approach philosophy: different methods can help us understand better. Just as science improves by using various research techniques, philosophy can also gain from using a variety of strategies. Therefore, it is not irrational to utilise film alongside more traditional methods to achieve the epistemic goals of philosophy, though it should not be used independently. Indeed, this combination can be superior to traditional methods alone, even if film, considered individually, might be a less reliable philosophical tool than established approaches. Our decision is not limited to choosing between a screwdriver and a coin; we can combine both tools. For example, there may be situations where using a coin is the easiest and most effective way to start the task. Once the screw is nearly in place, it becomes necessary to switch to a screwdriver to tighten the screw more securely (Livingston 2009, 57).

The acknowledgement that traditional philosophical methods could, in principle, accomplish all philosophical tasks raises another question about the value of alternative approaches. If traditional methods are sufficient, what might justify using additional tools? One response might be that while traditional methods are sufficient, they are not always optimal.

Different approaches can provide various ways of understanding. Historically, new methods in philosophy have led to new ideas. Similarly, film might offer new ways to express philosophical ideas and enable new forms of philosophical thought.

While helpful, this response to the rationality objection raises further questions about what makes film valuable as a philosophical tool. If film is to be more than just a pedagogical aid, we need to understand what distinctive contributions it might make to philosophy.

What abilities does film have that could help or even improve philosophical investigations? How can these abilities enhance traditional philosophical methods? This leads us to explore Livingston's bold thesis.

4.2.2 Bold thesis: the means condition and the results condition

Applied to the case of cinematic thought experiments, we may ask why we would choose to watch films as philosophical thought experiments instead of just reading them directly from philosophy books or articles. The rationality objection gains particular force when considered alongside Livingston's "bold thesis". This is an attempt to specify precisely what would constitute a genuine philosophical contribution through film, setting demanding criteria.

According to Livingston, a "bold thesis" consists of two significant elements:

(1) "a conception of which sorts of *exclusive* capacities of the cinematic medium [...] are said to make a special contribution to philosophy";

(2) "claims about the significance and independence of the latter contribution" (Livingston 2006, 11).

The first is what Livingston calls the *means condition,* since it concerns the means through which the film philosophised. It relates to the so-called *medium specificity*, that is, the idea that film has specific or unique abilities, such as juxtaposition, editing, selective focus of the camera, and correlations between soundtrack and image, among others. For Livingston, such abilities offer articulated content that is essentially non-linguistic but has a visual form (Livingston 2009, 12). This requirement demands that film's philosophical contribution emerge from capabilities unique to the medium itself, not merely from its ability to record or transmit philosophical content created through other means. The implications of this approach, particularly whether it risks confining cinematic philosophy to reflection on cinema itself rather than enabling broader philosophical potential, require closer examination, which we will return to in our revised account of the bold thesis. So, the *means condition* relates to the extent to which the cinematic work's epistemic property is realised through devices exclusive to cinema. This could be based on claims about cinema's exclusive essence. Livingston notes that judgments about cinematic specificity often stem from established practices rather than essential constraints. For example, while making a film consisting entirely of scrolling philosophical text is technically possible, this would be considered highly uncinematic relative to standard practices (Livingston 2019).

The second is the *results condition*. It suggests that the philosophical contribution made by film must be "historically innovative and independent" (Livingston 2009, 12). This requirement sets a high bar: film must present philosophy and make novel philosophical contributions that could not have been made through traditional means. The contribution must stand as an independent philosophical achievement, not merely an illustration or explanation of existing philosophical ideas. The results condition pertains to how effectively a cinematic work achieves what is referred to as the "epistemic philosophical desideratum," which represents the philosophical outcomes anticipated in an instance of cinematic philosophy (Livingston 2019, 81). This complex property includes several important factors: soundness, coherence, justification, novelty, generality, explanatory value, and relevance to specific philosophical issues. The extent to which a work possesses this epistemic quality can be viewed on a scale. However, we can generally classify works as having either high or low merit for initial analysis.

By combining these conditions – the *means* condition of cinematic specificity and the *results* condition of epistemic merit – Livingston identifies four categories of cinematic philosophy:

1. Low epistemic merit and low cinematic specificity (e.g. poorly made videos of mediocre philosophical lectures)
2. High epistemic merit but low cinematic specificity (e.g. excellent recorded philosophy lectures)

3. High cinematic specificity but low epistemic merit (e.g. philosophically simplistic ideas presented through sophisticated cinematic techniques)
4. High epistemic merit and high cinematic specificity (films where cinematic techniques are integral to philosophical investigation, such as films that use point-of-view shots, editing, and visual metaphors to explore epistemological or ethical questions) (Ibid., 86).

As Livingston frames it, the bold thesis requires films to achieve category 4: making innovative philosophical contributions through means exclusive to cinema. However, he argues this is likely impossible due to the "*dilemma of paraphrase*".

4.2.3 Dilemma of the paraphrase

Faced with the two elements required by a bold thesis, Livingston argues that a dilemma emerges, called the *dilemma of the paraphrase.*

The first horn of the dilemma presents us with a significant challenge: either a film's philosophical contribution can be paraphrased into language, in which case it fails the independence criterion and is not exclusively cinematic, or it cannot be paraphrased, which raises reasonable doubt about whether any genuine philosophical contribution has been made at all. Analysing or discussing a film's philosophical impact seems to compromise its assertion of independence (Livingston 2006, 12).

The second horn of the dilemma is that if we "accept [...] a more plausible, yet less narrow, conception of the cinema's exclusive capacities [this] leads to a trivialisation of the thesis that cinema can contribute to philosophy" (Livingston 2006, 12). This suggests that weakening the requirements to make them more achievable effectively undermines the bold thesis. If we broaden what counts as a film-specific contribution, we risk claiming that film can contribute to a trivially true but philosophically uninteresting philosophy.

So, in sum, cinematic philosophical ideas cannot be paraphrased, casting doubt on their existence, or they can be paraphrased, making them no longer exclusively cinematic.

Faced with this dilemma, Paisley Livingston offers three possible options:

a) to give up the exclusivity or specificity of the cinematic medium while maintaining the epistemic requirements;

b) to maintain the exclusivity or specificity of the medium, giving up the epistemic requirements;

c) to desist from both conditions, doubting something exclusive to the medium of cinema and reducing the epistemic importance of film to pedagogical and heuristic functions (Livingston 2006, 15-16).

These options represent different ways of resolving the tension between medium specificity and philosophical contribution. Each involves sacrificing some aspect of the original bold thesis, suggesting that the thesis as originally formulated may be unsustainable.

The philosopher thus opts for the third option (c), stating that "Films can be used to provide vivid and emotionally engaging illustrations of philosophical issues and ideas, and [...] sometimes yielding enhanced philosophical understanding" (Livingston 2009, 11). This conclusion effectively abandons the ambitious claims of the bold thesis in favour of a more modest role for film in philosophical discourse.

For some other philosophers, these parameters are unnecessary to defend film as philosophy. Livingston himself predicted that a strategy to maintain the FAP thesis would be to *weaken* the two conditions for a bold thesis, which would weaken the thesis itself. This prediction has proven accurate, as many contemporary defenders of film's philosophical potential have sought ways to preserve film's distinctive contribution while avoiding the rigid requirements of the bold thesis.

4.2.4 The means condition and medium specificity

To evaluate whether films can make genuine philosophical contributions through distinctively cinematic means, we must first understand what makes something "distinctively cinematic." The means condition demands that film utilise its specific or unique abilities to contribute to philosophy, abilities like juxtaposition, editing, selective focus, and correlations between soundtrack and image. However, determining an "exclusive" film capacity has been a central debate in film theory, with no definite answers. If we demand absolute uniqueness, we risk setting an impossible standard; if we accept any cinematic technique as sufficient, we risk trivialising film's philosophical contributions.

As discussed before, Mulhall sees the possibility of films themselves reflecting on and evaluating perspectives and arguments, thinking systematically and seriously like philosophers. For Julian Baggini, the path to arguing in favour of this possibility is to establish the difference between what can be said and what can be shown. According to Baggini, Mulhall's interpretation of the *Alien* films might be trying to show how cinema can do philosophy rather than tell how (Baggini 2003). In other words, for the philosopher, Mulhall "is showing us what philosophy is or can be instead of telling us, since his way of showing can be through demonstration or exemplification" (Ibid).

Nathan Anderson also suggests a defence of medium specificity as fundamental to defending film as philosophy: "What is interesting about cinema as a medium for 'doing philosophy' is the possibility that the way cinema 'philosophises' is distinct from the way it is done in texts or the classroom" (Anderson 2003). To assist Mulhall in demonstrating how film engages in philosophy, Anderson proposes that foundational theses about the nature of cinema are crucial in arguing that films can engage in philosophy. He refers to two essentialist theses regarding cinema: Eisenstein's formalism and Bazin's realism. This appeal to cinematic essentialism creates its own problems, as we will see.

The debate about medium specificity provides an important context for these discussions. Early theories of medium specificity often sought to identify film's essential nature. Greenberg's approach to medium specificity, though primarily focused on painting, influenced film theory.

Greenberg argued that each artistic medium should focus on what it alone can do, emphasising the importance of identifying and exploring a medium's unique properties. Greenberg argued that modernist art, particularly painting, should critically examine and emphasise its intrinsic properties, focusing on what makes each artistic medium unique (Greenberg, 1965). For painting, this meant an investigation of its fundamental characteristics, most notably its inherent flatness. Greenberg posited that modernist painters like Manet and the Impressionists began to foreground the two-dimensional surface, deliberately rejecting illusionistic techniques that had traditionally attempted to create depth and sculptural dimensionality. This approach was not about destroying artistic tradition but rather about "entrenching [the medium] more firmly in its area of competence" (Ibid) by stripping away effects from other art forms. The theory suggests that each art form recognises and utilises its distinctive qualities, with painting's defining feature being its flat surface.

Arnheim's theory of film medium specificity centres on understanding the unique characteristics that distinguish film from other arts. He argues that film is not simply a mechanical reproduction of reality but a distinct artistic medium with its own expressive potential. Arnheim's central idea is that film's limitations are its greatest artistic strengths. These limitations include the reduction of three-dimensional reality to a two-dimensional surface, the absence of colour (in early films), the fragmentation of continuous motion into discrete frames, and the absence of recorded sound (Ibid.). Arnheim emphasises that film's artistic power emerges precisely from its departure from natural perception. He explains that "the technical development of the motion picture will soon carry the mechanical imitation of nature to an extreme" (Arnheim 1957, 154). However, artistic potential lies in understanding and exploring the medium's unique properties. For instance, he discusses how film can manipulate perception using camera angles, lighting, and montage techniques.

A central aspect of Arnheim's theory is that film artists must deliberately engage with the medium's specific characteristics. He argues, "It is the task of the actors and the director to emphasise the expressive qualities of motion and thereby to define the character of the entire film as well as that of the single scene and the single shot" (Ibid., 182). The reduction of depth and colour becomes an artistic opportunity rather than a limitation. Arnheim observes that "The reduction of depth serves [...] to emphasise the perspective superposition of objects" (Ibid., 59). This means that what might seem like a technical constraint provides unique artistic possibilities for expression and meaning-making.

Moreover, Arnheim sees film as fundamentally different from theatre and photography. He argues that film is not just a mechanical recording but an art form that can "create new realities" (Ibid., 133). The specificity of film is also explored through its treatment of motion. Arnheim suggests that movement is not just a way to convey the narrative's events; it also carries significant expressive power: how objects and people move within the film frame becomes a crucial element of artistic expression.

This view of medium specificity questions the idea that film is just a mechanical reproduction. Instead, it sees film as a distinct art form with its own way of expressing ideas. Arnheim's theory on film is notable for its concept of "partial illusionism," where he suggests that the untrue aspects of film images help engage viewers. For Arnheim, a convincing illusion does not need to be complete in every detail because fundamental features are sufficient for our engagement. Significantly, Arnheim contends that increasing film's verisimilitude (through sound, colour, or three-dimensional effects) would degrade the cinematic aesthetic experience, which relies on accepting partial illusion and attending to the formal capacities of the medium. By working with the medium's potentials and limitations, artists can direct viewers' attention to formal qualities, encouraging observation that attends to the various layers visible in the film and a particular interest in its configurational properties (Moura 2011).

In contrast, André Bazin's realist theory focuses on film's ability to capture reality. He believes that photography and cinema represent a major advancement in art because they can authentically depict the world. A key idea in Bazin's theory is the concept of mechanical reproduction. As he explains,

> For the first time, between the originating object and its reproduction, only the instrumentality of a nonliving agent intervenes. For the first time, an image of the world is formed automatically, without the creative intervention of man (Bazin 1960, 7).

This mechanical process sets film apart from all previous artistic media. Bazin views this automatic reproduction as more than a technical procedure; he suggests that photography can transfer reality from the original subject to its reproduction. Unlike painting, which shows the artist's personal touch, photography offers an objective view that gives a credibility that other art forms lack. This means photography goes beyond traditional art in connecting with reality. Bazin even claims that film plays a role in creating reality.

The technological specificity of film is essential to Bazin's theory. He notes that only "a photographic lens can give us the kind of image of the object capable of satisfying the deep need humans have to substitute for it something more than a mere approximation" (Ibid). This technological uniqueness elevates film from mere representation to a form of metaphysical preservation.

Bazin's emphasis on film's unique capacity for capturing physical reality finds a contemporary echo in what has become known as the "Transparency Doctrine". Theorists like Kendall Walton and Roger Scruton argue that photography, and by extension film, does not merely represent reality but presents it directly – we see through photographs to their subjects (Moura 2011).

In summary, Bazin argued that the core of film is its ability to capture reality through photography, doing so automatically and without human interference. This sets it apart from other art forms.

Carroll provides a new way to understand what makes film special without relying on essentialist definitions. Rather than attempting to identify the essential properties of film, Carroll proposes a more flexible theory that acknowledges film's characteristic features while avoiding the limitations of strict medium

essentialism discussed throughout this section. Carroll argues that we should focus not on what is unique to film but on what is characteristic or typical. For Carroll, it is sufficient that "movie-made philosophy should be *recognisably* cinematic" (Carroll 2017, 273). This position acknowledges that techniques like montage, while not exclusive to cinema, are nonetheless paradigmatic of it. Carroll argues for an ontological theory of moving images that critically challenges traditional medium-essentialist approaches to film theory. To establish a more flexible yet rigorous framework for understanding film's distinctiveness, Carroll proposes five necessary conditions for moving images. These conditions avoid the pitfalls of medium essentialism while still providing clear criteria for what makes film a recognisable form.

1. Detached Display: Moving images constitute a visual representation that the viewer cannot spatially orient. The spectator experiences a fundamentally "alienated vision" (Carroll 1996, 62) where the depicted space is disconnected from their immediate physical environment;
2. Technical Movement Possibility: Moving images belong to a class of visual artefacts where the impression of movement is always technically possible, even in static works. This distinguishes film from traditional static visual media like paintings (Ibid., 62-66);
3. Performance Token Generation: Unlike theatrical performances, film performances are generated from templates (such as film prints or digital files) rather than interpretive processes; and related to this,
4. Non-Performative Artwork Status: Film performances are not considered autonomous artworks and cannot be aesthetically evaluated as independent artistic achievements (Ibid., 66-70).
5. Two-Dimensionality: Moving images are explicitly two-dimensional visual representations (Carroll 1996, 70-71).

By establishing these conditions, Carroll deliberately avoids medium-specificity essentialism. He argues that film should not be understood through a singular, defining medium-specific characteristic but rather through a more flexible, adaptive framework that can accommodate technological and artistic transformations.

Carroll supports "piecemeal theorising" (Ibid., 75) which involves looking at film structures and mechanisms without trying to create an all-encompassing theory. His work goes beyond boundaries of different media, suggesting that "moving images" include not just cinema but also other types of time-based visual media. The importance of Carroll's ideas is that they reject narrow definitions and embrace the idea of moving images as a flexible and evolving form of art. This allows for inclusion of current and future ways of doing film. For example, we currently have 3D film which, apparently, would seem to counter the two-dimensionality criterion. However, two-dimensionality refers to the flat surface of the screen rather than the illusion of depth. 3D films, despite their perceptual effects, remain projected onto a two-dimensional plane and utilise representational techniques to simulate spatial depth. Given Carroll's methodological preference

for piecemeal theorising, it remains open to adaptation and could potentially accommodate developments such as stereoscopic cinema.

More recent technologies, like virtual reality (VR) with platforms like Oculus, are more challenging to the way we think about two-dimensional media. Instead of just watching a flat screen, VR allows users to step into a fully immersive three-dimensional space. This experience blurs the lines between the viewer and what they see, changing our relationship with movies. In VR, how we move and perceive things becomes essential to experiencing the moving image. This shift suggests we might need to rethink how these immersive technologies relate to Carroll's theory. However, instead of contradicting his ideas, these developments show that piecemeal theorising remains relevant because it can adapt to changes in technology without needing a complete overhaul of existing concepts. This collapses the traditional boundaries between viewer and image, disrupting the "detached display" that Carroll associates with cinematic spectatorship. Now, elements like spatial orientation, how we move around, and our sensory feedback are key parts of experiencing moving images. However, rather than undermining it, these developments illustrate the ongoing relevance of piecemeal theorising since it is an approach capable of adjusting to technological transformations in the moving image without requiring a complete revision of all concepts.

The discussions about what makes film unique help us understand how films can contribute to philosophy. Carroll suggests that instead of looking for what makes film special in a strict sense, we should focus on common features that films share. This way, we can explore how typical cinematic techniques can have philosophical meanings without falling into the trap of needing films to be completely unique.

4.2.5 The results condition

The results condition (requiring philosophical contributions to be "historically innovative and independent") creates what Livingston calls the dilemma of paraphrase. This condition presents a particularly difficult challenge for film's philosophical potential, as it demands that films make novel and independent contributions to philosophical discourse.

The dilemma emerges from the impossibility of simultaneously satisfying both aspects of the results condition. If a film's philosophical contribution can be paraphrased in traditional philosophical language, it seems to lose its claim to independence. If it cannot be paraphrased, its status as a philosophical contribution becomes questionable.

Carroll's response to Livingston's demands for originality and independence is twofold. First, he questions whether absolute originality is possible in philosophy, noting that "Little philosophy, if any, could survive such a test" (Carroll 2017, 274). Second, he suggests that a philosophical contribution can come from new ways of looking at existing ideas instead of just new concepts. He reminds the reader of Heidegger's exploration of human mortality, which serves as a reminder of important truths we often

overlook. This perspective helps us see that film contributes to philosophy not by being completely new, but by presenting and exploring philosophical ideas through its own means.

This response to Livingston highlights important ideas about what makes a contribution to philosophy. It suggests that originality might be more about offering new viewpoints, arguments, examples, and analyses of existing concepts rather than inventing completely new ideas. Second, it indicates that the value of philosophical work might lie in its novelty and its ability to present ideas in ways that make them newly accessible or meaningful.

While Carroll's response helps us understand why absolute originality should not be required for philosophical contribution, it does not fully resolve how films can make genuine philosophical arguments through specifically cinematic means. This brings us to more recent attempts to solve the dilemma of paraphrasing while preserving the film's distinctive philosophical potential.

4.2.6 Solving the dilemma

Carroll tries to solve the dilemma of paraphrase with a simple answer: in philosophy it is common practice to paraphrase philosophers. He notes that we regularly paraphrase Kant's work without diminishing its status as original philosophy. "But doesn't all philosophizing require a preexisting discursive context?" Carroll asks (Carroll 2017, 274). In film as philosophy, we must consider that paraphrasing occurs between two distinct media. In philosophy, we often paraphrase a philosopher's words into alternative phrasing, or we may convert text into formal logic and vice versa – all of which remain in written form. However, there is no direct transition between images and text or from text to images.

Several philosophers have attempted to navigate these challenges through modified versions of the bold thesis.

Cox and Levine propose a "moderate thesis" suggesting that "certain philosophical things are better done in film than in written texts" while acknowledging that this "would not require film to have unique access to its mode of philosophising" (Cox and Levine 2012, 10). This moderate position maintains that films can make valuable philosophical contributions through their specific techniques without proving absolute uniqueness or independence.

As Nathan Andersen has suggested, Mulhall rejects the need to accept a Bazinian realism or other essentialist cinema theory to accept that film can do philosophy. The philosopher says that he has paid attention to the technique of cinema in the way it allows certain films to make philosophy, referring to the role of Sarah Connor's photography in *Terminator*, Fincher's edition of *Alien*, or the angle in *Jeunet*. He argues that although other art forms can use these techniques, this does not imply that in cinema, its use does not allow it to do philosophy. Thus, for Mulhall, using exclusively cinematographic resources is not essential for the film to philosophise.

Aaron Smuts develops this line of thinking further, arguing that cinematic means that "include montage, camera angles and movement, and the juxtaposition of word and image" (Smuts 2009, 410), while not exclusive to cinema, are characteristic enough of it to fulfil the means condition.

To demonstrate how films can provide such reasons through specifically cinematic means, Smuts analyses Sergei Eisenstein's "For God and Country" sequence in *October* (1928) as an example of philosophical argumentation through montage. The sequence presents an analogical argument comparing Christian religious artefacts with "primitive" pagan statuary, suggesting through careful visual juxtaposition that both arise from similar motivations and are equally questionable. What makes this philosophically significant is the conclusion and the way the film provides reasons to believe it through specifically cinematic techniques.

The sequence begins with familiar Christian artefacts (cathedrals, statuary, artworks) and gradually intersperses these with increasingly "primitive" religious icons. Through montage, Eisenstein creates what Smuts describes as "a procession of images that traces a path through time" (Smuts 2009, 416), leading viewers to recognise similarities between the two categories of religious artefacts. Smuts argues, "Even if one rejects my claim that the sequence presents an analogical argument, if one is led to see some similarity between the two classes of artifacts, the sequence certainly provides a reason to believe its conclusion" (Ibid). This example helps address the dilemma of paraphrase in several ways. As Smuts argues:

> Yes, it is true that I cannot state the argument without using language, but this does not mean that the philosophical work done is exhausted by my statement. Through the means of montage the film articulates its analogical argument – that is, it provides reasons to believe its conclusion. Although the support can be adequately described, it cannot be completely reduced to language. Any mere description will lack the force of the original, nonlinguistic presentation (Ibid).

So, while the philosophical content can be stated linguistically, the force of the argument depends on its specifically cinematic presentation through montage. Also, the fact that we can paraphrase the argument does not mean we are its source, just as being able to paraphrase Nietzsche's arguments does not make us their author. The philosophical work is being done through cinematic means, even if we need language to discuss it.

Smuts provides another illuminating example through his analysis of *The Twilight Zone* William Claxton' episode "The Little People" (1962), which explores philosophical questions about worship and power through narrative and visual means. The episode argues that mere power, even unlimited power, is insufficient to make something worthy of worship. It does this not through explicit philosophical discourse but through the specifically cinematic tools of narrative, point of view shots, and visual scale contrasts between characters.

These examples suggest that while films may not be able to present formal logical arguments in the traditional sense, they can provide novel reasons to believe philosophical claims through their distinctive cinematic means. This allows us to maintain high standards for what counts as philosophical contribution while acknowledging film's capacity to do genuine philosophical work in ways that exploit its particular strengths as a medium.

With this understanding, we can now resolve the tension that emerges from Livingston's condition. Rather than requiring absolute originality or complete independence from linguistic expression, we can recognise films as making genuine philosophical contributions when they provide new reasons or perspectives on philosophical questions through their characteristic modes of presentation. Smuts notes, "[a]lthough there may have never been a film that has made an innovative and independent philosophical contribution, and it is certain that few have, there is no reason to think that films cannot do philosophy" (Ibid., 418-19).

Moreover, the dilemma rests on questionable assumptions about the relationship between form and content in philosophical thinking. The assumption that paraphrasability negates independence seems to ignore how philosophical ideas often depend on their specific mode of expression. Just as certain philosophical arguments might be most effectively presented through formal logic while others benefit from narrative exposition, some philosophical insights *might* be best conveyed through cinematic means. As we will see in later film analyses, questions about morality, perceptual manipulation, and the nature of fictional narratives are particularly amenable to cinematic investigation.

This understanding suggests another approach to the results condition. Rather than requiring absolute independence or complete novelty, we might consider how films can make distinctive contributions to philosophical understanding through their characteristic modes of presentation. This approach acknowledges that while a film's philosophical content might be paraphraseable, something significant about its philosophical work might be "lost in translation".

The implications of this revised understanding go beyond merely defending the philosophical potential of film. They indicate a need for us to reconsider our criteria for what counts as a philosophical contribution in a broader sense. This discussion will be addressed in a later chapter. Suppose we accept Carroll's argument that even traditional philosophy rarely meets absolute originality or independence standards. In that case, we might better appreciate how different modes of philosophical expression, including film, can contribute to philosophical understanding in distinctive ways.

4.2.7 A revised bold thesis

We began by asking whether films could present thought experiments with the same or analogous value as written ones. An answer to this question can now be offered by examining the rationality objection, bold thesis requirements, and various theoretical responses. Films can make philosophical contributions, but not

because they possess unique capabilities or complete independence from linguistic expression. Instead, their philosophical potential lies in their ability to deploy characteristically cinematic techniques to provide new reasons for believing philosophical claims or new ways of understanding philosophical problems.

Digital technologies challenge medium essentialism. The proliferation of hybrid media forms challenges traditional notions of what constitutes "cinema." Focusing on pure medium specificity becomes more problematic when films regularly incorporate elements from other media forms, including social media, video games, and interactive elements. These hybrid forms do not diminish film's philosophical potential. This shows how the means condition does not depend on medium specificity but on identifying film's main features that are commonly identified.

The examples Smuts examined may demonstrate how films can philosophise. While these contributions can be paraphrased linguistically, such paraphrases do not exhaust or replace the philosophical work done through cinematic means. Davies, too, proposes that:

> For film to be capable of providing philosophical insights that are medium exclusive relative to the verbal medium employed in sound cinema, there must be some film *F* and some philosophically relevant content PC_a such that *F* articulates PC_a while it is not the case that PC_a is articulated by the verbal content of *F*" (Davies 2019, 142)

This formulation allows for philosophical content to emerge from the film as a whole, including how visual elements might provide "a critical commentary on what is being said, so that the overall philosophically relevant content articulated by the film differs significantly from the philosophically relevant content articulated verbally" (Ibid., 143). So, while we can paraphrase the philosophical content present in the film, this content is not articulated verbally in the film, or we should say, at least not directly verbally. If a film is entirely composed of a philosophical dialogue, this does not count as an instance of film as philosophy, as this dialogue could be completely put into text, and it would not change its philosophical significance, as the film as a medium did not contribute at all.

This understanding helps us move beyond the rationality objection and the dilemma of paraphrase. Films do not need to be the most efficient or unique tools for philosophy to be a valuable contributor to philosophy. Instead, they should offer different, but complementary approaches that can enrich philosophical understanding through their characteristic modes of expression. This suggests that the relationship between film and philosophy is not best understood through an either/or dichotomy but rather through consideration of how different media can contribute distinctively to philosophy.

It also suggests that we must clearly distinguish between different types of relationships between cinema and philosophy. When we defend that films can function as philosophical thought experiments, we are not asserting that any film containing philosophical dialogue is automatically philosophical as film.

Our moderate thesis requires that specifically cinematic elements contribute to the thought experiment. A film composed exclusively of philosophical dialogue fails this, as removing the visual elements and retaining only the audio would not result in significant philosophical loss. A film without dialogue can still potentially philosophise; a film consisting only of dialogue does nothing special. In such cases, the film itself is irrelevant. Furthermore, the mere footage of philosophical lectures, for instance, such as a video of a conference or a YouTube video explaining certain philosophical ideas, contributes nothing to the philosophy being done.

This distinction becomes clearer if we consider three possible categories of cinematic use in philosophy:

a) Film as transmitters: Here, film functions merely as a recording and distribution technology. A filmed philosophical lecture retains all its philosophical value when transcribed. The cinematic medium is philosophically irrelevant: we could achieve the same result with audio or text.
b) Film as popularisation: Film spreads philosophical ideas in more accessible or memorable ways, but without adding new philosophical content. Here emerges an interesting question about popularisation: does it have epistemic value in its own right? While we recognise the potential epistemic value and especially the social value of disseminating philosophical ideas through cinematic means, this may not say much about philosophical value specifically, and this is not our primary concern.
c) Film as a philosophical medium: The philosophical experience emerges through cinematic specificity.

We are interested in how cinema can actively participate in the production of philosophical knowledge, not merely in its transmission or popularisation

Having established that films may make philosophical contributions through their characteristic techniques without requiring medium specificity, absolute uniqueness or complete independence from linguistic expression, let us now examine how films may be viewed as philosophical thought experiments

4.3 Cinematic philosophy through thought experimentation

In this section, I will be making arguments for the thesis that films can be philosophical thought experiments, and to do so, I will also be addressing some of the main objections made against thought experiments in general and against artistic and cinematic thought experiments in particular.

McClelland identifies several key challenges that any defence of "film as philosophical thought experiment" must address.

First is the question of how we can justify regarding a cinematic thought experiment as specifically philosophical rather than scientific or belonging to another discipline: "Granting that some films can act as thought experiments, how (if at all) can we justify regarding them as *philosophical* thought experiments?" (McClelland 2019, 93).

As discussed in the previous chapter, thought experiments are deployed across various fields, including physics, biology, economics, history and mathematics. While philosophers such as Wartenberg have recognised that thought experiments are not the exclusive domain of the philosopher, there has been insufficient attention to what makes a cinematic thought experiment distinctively philosophical. For this, we can use Wartenberg's three conceptions of philosophy we should regard inclusively:

(1) Philosophy as approaching a set of "eternal questions," such as knowing what a moral action is (a topical account);

(2) Philosophy as a meta-discipline, i.e., as a discipline that asks questions about the nature of other disciplines, such as sciences and arts;

(3) Philosophy as a set of methods, as the use of explicit argumentation and formal logic, or the use of thought experiments (a methodological account);

Wartenberg argues that we should precisely regard these conceptions in a "mutually supportive" way to avoid excluding or including too much in a conception of philosophy (Wartenberg 2007, 29-31). So, we should take into account that a work either being about a philosophical subject, or employing a philosophical method by itself is not enough to qualify it as philosophy. One approach is to characterise philosophy by its subject matter. Philosophy writings generally engage with questions about "eternal subjects" like ethics, knowledge and the nature of reality by offering explicit argumentation. So, if other works engage with those same issues, they should qualify as philosophical in nature. As McClelland affirms, "a work contributes to philosophy when it contributes to our philosophical knowledge" and "knowledge is philosophical when it is knowledge of a philosophical subject" (McClelland 2019, 97). This approach of characterising philosophy by subject matter may appear circular. What makes a subject philosophical in the first place? As McClelland notes, it may be that such subjects are "eternal", i.e., they are related to fundamental human concerns, and may resist definite resolution.

However, this does not mean they are philosophy themselves yet, as this only would match one of the three conceptions. We need to argue that a film should not only engage with one conception of philosophy but several. It should be philosophical, meaning its subject matter should address "eternal questions," and it should employ a philosophical method, specifically, by functioning as a thought experiment.

McClelland's strategy is to suggest that film may offer unique advantages for philosophical thought experiments, over traditional ones. He proposes several ways in which the cinematic medium enhances

rather than compromises philosophical inquiry through thought experiments. He does this by presenting some sceptical views of thought experiments in philosophy and then arguing how cinematic thought experiments may address these doubts (McClelland 2019).

Those and other sceptical views will be addressed, some of which were previously analysed.

4.3.1 The value of thought experiments in philosophy

Firstly, both advocates and critics of FAP sometimes support that films may, indeed, be seen as thought experiments, but that is not relevant enough to think that films can philosophise in a relevant manner. Bruce Russell admits that films can be counterexamples to philosophical theses. However, he does not consider this an important philosophical contribution, as it is limited by its inability to provide the explicit argumentation that Russell considers essential to genuine philosophical work (Russell 2001, 166). We should recognise that, although it may seem like a limited contribution, providing a counterexample to a philosophical thesis is a valid contribution to the philosophical discussion at hand, so we should not rush to downplay its value. Moreover, cinematic counterexamples might actually be more effective than traditional philosophical thought experiments in this regard, as will be argued.

Richard Eldridge, for example, also downplays the potential power of cinematic thought experiments. Although he supports film as philosophy, he believes that an argument for cinematic thought experiments is uninteresting since, in philosophy, they are minor projects: "I am inclined to think that proof, thought experiment, and counterexample are all less important in successful efforts to think more clearly than is often supposed. So it is, for me, somewhat beside the point to say that film can do those things." (Eldridge 2009, 113) However, as Wartenberg points out, although these tools are a fraction of what philosophy does, several are central to specific philosophical ideas, such as the Evil Genius in Descartes' *Meditations* and the Ring of Gyges in the *Republic*. Thus, for the philosopher, "we need to emphasise the importance and vivacity of such thought experiments in philosophy proper to give the thesis that films contain philosophical thought experiments more punch and greater plausibility" (Wartenberg 2011, 21). Considering thought experiments as "less important" in philosophy can be problematic. Even if they may be small parts in philosophical works, in terms of their length, they seem to have large roles in them, as seen in the previous chapter – to the point where a lot of philosophy is done about the value of thought experiments in philosophy (and in science).

Finally, if we are sceptical towards the value of thought experiments in general, then even if we argue that films can be thought experiments, that does not amount to anything of value for philosophy. However, this does not have to be the case. Most problems directed against thought experiments are not so much definite objections against the use of thought experiments overall but can be interpreted as suggestions for how they can be better. And one way of trying to make thought experiments better is to find cinematic ones that are able to respond to those concerns. This is exactly the strategy used by Tom

McClelland when he frames some challenges for thought experiments as opportunities for cinematic ones (2019). This is also a strategy I will be implementing and expanding on.

4.3.2 The problem of ambiguity, and the "Socratic model"

One way of calling into question the idea that cinematic thought experiments are ways by which one can argue in favour of FAP is to make irreconcilable distinctions between philosophical thought experiments and cinematic fictional narratives. For Wartenberg, this kind of distinction and a priori argument does not solve whether films can philosophise, since, to know if the possibility of cinematic thought experiments is feasible, we should debate specific interpretations (Wartenberg 2009, 121). However, I should address those worries.

Deborah Knight has articulated several distinctions between philosophical thought experiments and artistic fictional narratives, which can help us explore the sceptical position. Knight's critique of Carroll's thesis that *The Third Man* functions as a philosophical thought experiment on friendship provides a helpful starting point.

One important difference she points out is that philosophical thought experiments tend not to be resolved, and fictional narratives are usually resolved. Philosophical thought experiments are usually open-ended since "their whole point is to challenge our intuitions about given situations by illustrating how difficult it is to be clear about certain concepts, and to provoke further debate" (Knight 2011, 289). Fictional narratives, such as literary narratives, have a resolution, or closed end, and their purpose, unlike philosophical thought experiments, is to think "about such things as narrative structure and genre, neither of which is relevant to philosophical thought experiments" (Knight 2011, 289).

First, let us recognise that philosophical thought experiments can indeed take genre and narrative structure as their subject matter. This can be present in philosophical thought experiments that analyse narrative structures and genre conventions, as well as the structures of thought experiments themselves. For example, in *The Philosophy of Horror* (1990) to examine the horror genre and its most common narrative structures, Carroll refers to many horror books and films, explaining their structure – and arguing that the most common one is the "complex discovery plot" (Carroll 1990, 99). But as he refers to actual films, we can as easily imagine hypothetical scenarios, like hypothetical films, in order to analyse their genre and their narrative structure.

There are also philosophical thought experiments that are concerned with their own structure – meta-thought experiments, we could say. For example, in *Intuition Pumps and Other Tools for Thinking* (2013), Dennett dissects the p-zombie thought experiment by imagining an alternative "zimbo" called Zeke. While p-zombies are physically identical to humans but lack consciousness, Dennett's zimbo possesses self-monitoring capabilities. Zeke has internal systems that not only processes information from the environment but can also monitor the processing activities and itself. Though still technically unconscious, those self-

monitoring capabilities make zimboes functionally similar to conscious humans (Dennett 2013, chapter 55). Dennett does not merely use this scenario to discuss consciousness; he analyses how the structure of this "intuition pump" (the p-zombie and the zimbo) works, showing how its narrative construction guides our intuitions in particular directions. Similarly, by analysing Mary, the colour scientist, he demonstrates how variations of the scenario's structure shift our intuitions drastically (Ibid., chapter 64). These philosophical thought experiments and their variations have the explicit aim of exploring the narrative structure of thought experiments themselves.

So, as we can see, there are philosophical thought experiments that are indeed preoccupied with exploring genre and narrative structures. It is also interesting that Knight attributes these themes as the main aim of artistic works, as there can be films that are not just implicitly, but explicitly and directly concerned about exactly that. Structural films that emerged in the 1960s as experimental films are examples of this. They were concerned with demystifying the filmmaking process, revealing their fundamental and structural elements. They directly rejected narrative as essential for film. And Carroll argues that *Serene Velocity*, for instance, is exactly this type of film that is a thought experiment about the nature of film itself – we could say, in Knight's words, its narrative structure and genre. In such a case, films are not just generally preoccupied with employing narrative structures belonging to their own genre, like a horror film would be preoccupied with employing common plots found in the horror genre; they actually question that entirely.

It is not true that all philosophical thought experiments are open-ended or that all films are not, either. One of Wilson's critiques against thought experiments in philosophy is precisely that they are biased and constrain interpretation. Also, some films are famously open-ended. For example, *Blade Runner* leaves Deckard's identity an open question. However, if we consider the difference Knight poses at face value, we can connect it to the explicitness objection. Any thought experiment or example requires an explicit argument that connects it to a general philosophical conclusion. As Russell states, "thought experiments can be a stimulus to argumentation, can motivate the listener, reader, or viewer to produce an argument, but they are not arguments themselves" (Russell 2007, 13). Even if we agreed with Norton's argument view, a thought experiment would have to be reconstructed as an argument. So, if thought experiments were always open-ended, they would need to be accompanied by explicit argumentation. Now, Carroll responds to this by stating that although a thought experiment in not an explicit argument, it is a device that, although sometimes inconclusive, is used in the same way as other similar devices, such as a rhetorical question, in which the function of clarifying or explicitly reasoning about the thought experiments is attributed to the reader (or spectator) (Carroll 2008, 4).

It is quite ironic to observe that one of the characteristics often attributed to films is their ambiguity and openness to interpretation. If we embrace Knight's perspective that philosophical thought experiments are similarly open-ended, we could argue that they too may be criticised for being overly ambiguous and

subject to interpretation. We may interpret Knight's view of the open-ended nature of thought experiments as being a virtue, as it is this open-endedness that creates a lot of interesting philosophical debates. And this is an argument that we could make for an advantage of cinematic thought experiments: their actual ambiguity may be advantageous. If we apply the same concerns about philosophical thought experiments to films, we can argue that when the content of a cinematic thought experiment is open to interpretation, rather than restrictive, that is an epistemic virtue. This is an argument that McClelland makes for the case of film as philosophy in general, and that we can apply for cinematic thought experiments in particular. He argues for a "Socratic Model", where what Carroll notes is what happens: the function of explicitly reasoning about the thought experiment, or the film, is attributed to the spectator. McClelland notes that some cases of philosophy do not reach firm conclusions at all but rather enhance understanding by inviting reflection on one's own beliefs. As he notes,

> If Frank Jackson had just said 'here's an intriguing scenario about a vision scientist in an achromatic prison – now what should we make of it?' his paper would still have been a valuable contribution to the philosophy of mind. In fact, Jackson's TE has been more influential than the conclusion he originally extrapolated from it (McClelland 2019, 96).

As he proposed for the Socratic Model, while films may not philosophise in the traditional sense – by articulating and defending precise arguments – they can do so by stimulating reflection in their audiences (McClelland 2011, 6-7). This model draws its name from the Socratic method of not directly communicating knowledge, but by guiding interlocutors through questioning and discovering philosophical insights for themselves. Wartenberg also argues that Cora Diamond demonstrates how philosophy can be conveyed through fictional narratives by illustrating this with the example of Socrates engaging in dialogue with his friends about the dilemma of choosing between facing condemnation or fleeing Athens (Diamond *apud* Wartenberg 2016, 170). This narrative, whether fictional or not, encourages people to perceive the world in a different, unfamiliar way.

Thus, in the case of cinematic thought experiments, they can present scenarios that invite viewers to engage in philosophical inquiry, thus contributing to the broader philosophical conversation without claiming full autonomy. This model addresses both the explicitness objection, or the problem of ambiguity, and the generality objection. In this case, films are not to be viewed in isolation, as autonomous and independent philosophical pieces. We will see more on this exact point further. So, if we follow a Socratic model of viewing cinematic thought experiments, we may actually recognise their advantage of not biasing the viewers towards specific conclusion. As McClelland argues, "the key lesson of Socrates' slave boy scenario is that following another person's reasoning might give one 'true opinion,' but only by reasoning to the conclusion for ourselves do we achieve genuine knowledge" and "if we are left to reason for ourselves there is less risk of us being misled by the author of the TE." (McClelland 2019, 104).

In sum, Knight argues that films are resolved, contrary to philosophical thought experiments that are open-ended. But this is false, though. In fact, we could argue that films' supposed lack of explicit argumentation, the target of the explicitness objection, actually supports their open-ended nature. To address this, we can argue for the Socratic model, by recognising that knowledge derives not from explicit argumentation but through the audience's engagement with the film, an advantage over classic thought experiments that may bias us towards a predetermined conclusion.

4.3.3 The issue of detail

Another difference that Deborah Knight points out is that in philosophical thought experiments, the narrative aspects are not essential to making the conceptual point, whereas in artistic works, story and characters are what we value (Knight 2011, 280). Carroll challenges this by questioning how philosophical thought experiments could promote thinking if their narrative elements were unnecessary: "But if thought is reached by means of the story that embodies the thought experiment, why is it unnecessary?" (Carroll 2017, 281).

Indeed, detail is significant for philosophical thought experiments. As Dennett showed with his variations of famous thought experiments, the details matter because they shape the intuition pump differently, causing different intuitions in the reader; they do influence interpretations and arguments that can be made as a result.

Thus, the fact that films provide way more detailed scenarios may actually be an advantage. As McClelland observes, "besides being more detailed in general, film also has the advantage that some potentially important details are much easier to depict than textually" (McClelland 2019, 96).

The issue of detail is central to Wilkes' critique of thought experiments. Wilkes argues that they lack the necessary background conditions. McClelland responds to this by acknowledging that the lack of detail can, indeed, put the epistemic value of thought experiments in question. And, thus, he argues that this is why cinematic thought experiments may have an advantage over textual ones. Dancy also argues that thought experiments have limited utility due to their incompleteness and a lack of specification, in contrast to actual cases that are far more detailed ("richer in properties").

Films present long and complex narratives, offering the rich contextual background that may lack in traditional thought experiments. As McClelland states, "If we're considering the film as whole to constitute a TE, it will have more detail than almost any academic TE" (McClelland 2019, 101). He gives the example of *Three Billboards Outside Ebbing Misouri* (2017), directed by Martin McDonagh, as presenting ethical questions about justice and revenge through intricate details: "The life that the protagonist and her family lead after the murder of her daughter, the suicide of the police chief that oversaw the investigation and the redemption of the racist officer who eventually acts in support of the protagonist all contribute to the moral lessons we draw from the film (Ibid., 101). This is the kind of rich detail that abstract philosophical texts cannot match. The narrative complexity allows viewers to grasp how moral principles operate in concrete

circumstances, revealing tensions that abstract argumentation might simplify away. This simple explanation of the elements that convey the moral lessons of the film is lacklustre compared to watching the film. This aligns with Davies' claim that the rich narrative detail in films like *Blade Runner* "provides greater insight into what is at stake in philosophical debates" (Davies 2019, 140).

In his analysis of *Blade Runner*, Davies demonstrates how the detailed narrative gives a richer exploration of personhood, in contrast to Putnam's abstract thought experiment on artificial consciousness.

Putnam explored the question of whether robots could be considered conscious. His main argument is that determining whether it is not a matter of discovery, but of decision, that is, we must decide if we want to extend our concept of consciousness to include robots. To establish this, he makes a thought experiment where there is a robot called Oscar who is "psychologically isomorphic" (Putnam 1964, 677) to humans, so Oscar obeys the same psychological laws and exhibits the same kinds of behaviours that humans do: "Let Oscar be one of these robots, and let us imagine that Oscar is having the 'sensation' of red. Is Oscar having the sensation of red? In more ordinary language: is Oscar *seeing* anything? Is he thinking, feeling anything? Is Oscar Alive? Is Oscar Conscious?" (Ibid., 678) He then imagines robots creating their own robots (which he calls ROBOTS). Just as we question whether robots have consciousness, they may also question whether their own ROBOTS have consciousness and, thus, he argues that "our position with respect to robots is *exactly* that of robots with respect to ROBOTS. There is not the slightest reason for us, either, to believe that 'consciousness' is a well-defined property, which each robot either *has* or *lacks*" (Ibid., 689).

According to Davies, Putnam's paper "in no way explores or clarifies what this involves and what import it has" (Davies 2008, 7). By contrast, *Blade Runner* illuminates what it actually means to treat something or someone as conscious, by presenting us with "replicants", androids that are almost perfect replicas of humans[11]: "to admit it to one's moral community, to hold it responsible for its conduct, to feel both with it and for it" (Ibid). These come through in the film, but "are not clear in the dry presentation of Putnam's TE" (Ibid).



[11] In the film, the only way of knowing that we are faced with a replicant instead of a human is to do the Voight-Kampff test that measures bodily functions such as respiration, heart rate, blushing and pupillary dilation in response to emotional questions.

Figure 11: Still image of Roy Batty, the leader replicant responsible for escaping a group of replicants. From *Blade Runner* (directed by Ridley Scott, Warner Bros., 1982)

For Davies, the value of all these details is that we gain understanding of the philosophical concepts through dramatic imagining in our engagement with film as "a visually and aurally present manifold" (Davies 2019, 143), not afterwards when we abstract by paraphrasing the cinematic details.

Both McClelland and Davies argue that film's rich details enhance the philosophical thought experiment, rather than being a problem. In this aspect, films may have an advantage over textual philosophical thought experiments, if only for the fact that they are much more detailed. As previously mentioned, Wilkes critiques Plato's Ring of Gyges because it fails to specify important background conditions, like knowing if the owner of the ring becomes intangible, other than invisible, and this impedes us from evaluating the scenario properly. By contrast, a film would definitely establish such background conditions in full detail, as even in science fiction, films operate under an internal logic that is explained by their own details. For example, we can imagine that if Ring of Gyges were to be portrayed in a film, it would have to address Wilkes' concerns. How would the owner of the ring walk around a building, for instance? Would it open doors, or cross them? A film would portray all of those facts. There are films that actually do tackle invisibility. For example, *The Invisible Man* by Leigh Whannell (2020) establishes the logic for how invisibility works in it. In the film, it is achieved through a technologically advanced suit lined with optical cameras that bend light around the person wearing it. The suit can malfunction and produce subtle distortions when moving, and it does not mask sound. These details here are not peripheral either, as they establish how the protagonist deals with her tormenter wearing the suit. The film also addresses how this "invisible" person navigates physical spaces and interacts with objects. Now, this does not mean that this film, or any other, philosophises just because it is rich in detail. We are just addressing the fact that films may offer better background conditions, through their details, to construct effective philosophical thought experiments.

4.3.4 The role of emotion and the "proper domains" of psychological capacities

Another difference that Knight finds between thought experiments and film is that "philosophical thought experiments do not rely on our emotional engagement or our desire for some narrative outcome rather than another" (Knight 2011, 289), whereas films do rely on our emotional engagement. For Carroll, this is "flat-out false" (Carroll 2017, 282), since many thought experiments do provoke and, we may say, rely on emotion, especially in ethics and political philosophy, like Thomson's Violinist. As Carroll states, "at the very least, the emotions these thought experiments evoke are germane to our reflection upon these thought

experiments and, in some cases, may be arguably decisive" (Ibid). This is because our intuitions are connected to our emotions.

Even if we argued that philosophical thought experiments do not rely on our emotional engagement, emotional engagement does not need to be a problem for them, quite the contrary. We can actually argue that the affective dimension of cinema provides another advantage for thought experimentation. Cox and Levine, for instance, argued that films can engage both rational and emotional faculties simultaneously: "some films, for some people at some times, are able to generate the kind of attentiveness and emotional insight that may undermine prior commitments and suppositions even when they had been thought to be intellectually and rationally grounded" (Cox and Levine 2012, 17). This ability that film has to engage us emotionally can be valuable to thought experiments where our moral intuitions are to be challenged.

Davies also argues that emotional engagement in cinematic thought experiments is not a distraction from philosophical clarity, but the affective dimensions of film enable forms of understanding that are not accessible only via abstract textual thought experiments. When he analyses *Blade Runner*, Davies emphasises that the film's philosophical contribution depends on its ability to engage viewers emotionally with the plight of the replicants.

He also analyses *The Thin Red Line*, arguing that the film's "subjective perspective, impressionistic images and sounds and an episodic narrative" generate an "overall emotional or affective experience" that is "primarily perceptual and embodied rather than cognitive and evaluative" (Davies 2019, 148) Here, film's aesthetic dimensions include its affective impact, and, as Davies argues, we "come to reflectively experience the embodied nature of the characters' engagement with their lived worlds" (Ibid., 151).

Martha Nussbaum also argued that emotions have a role in narrative understanding. She challenges the traditional philosophical separation between emotion and cognition, by arguing that emotions "are not simply blind surges of affect, recognized, and discriminated from one another, by their felt quality alone; rather they are discriminating responses closely connected with beliefs about how things are and what is important" (Nussbaum 1990, 41). Thus, emotions are intelligent responses integral to human reasoning (Ibid). Gendler also supports this perspective in the value of emotion in thought experimentation. She also acknowledges that the traditional division between emotional and rational processes is incorrect, drawing on Damasio's findings on the role of emotion in reasoning. As such, she argues that thought experiments activate "cognitive mechanisms associated with vivid imagining" that integrate emotional and rational processing, leading subjects to potentially "reverse a prior commitment" in ways that abstract reasoning might not achieve (Gendler 2007, 69). This connects, as Gendler argues, to Damasio's work demonstrating that patients with damage to emotional processing centres of the brain often exhibit impairment in decision-making processes (Ibid., 77).

This could suggest that a film's ability to provoke emotional responses is not a distraction, but a necessary aspect. An illustration of this can be found by comparing the standard philosophical presentation

of the trolley problem with its depiction in the TV series *The Good Place*. In text, the trolley problem is presented as an abstract scenario: a runaway trolley is headed toward five people who will be killed unless you pull a lever to divert it to another track where it will kill one person. This presentation strips away emotional context to isolate the moral principles at stake. By contrast, *The Good Place* transforms this thought experiment into a visceral simulation where characters are repeatedly placed in increasingly graphic and emotionally charged trolley scenarios, complete with realistic screams, blood splatter, and traumatic consequences.

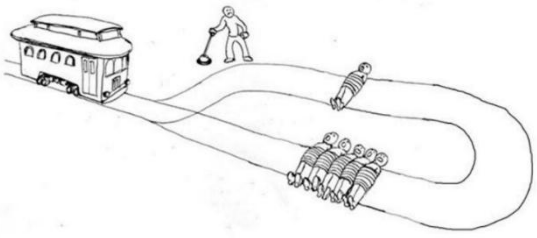


Figure 12: Simplified illustration showing the classic trolley dilemma with a track switch.



Figure 13: Still image from *The Good Place* (Season 2, Episode 5, "The Trolley Problem") showing Chidi experiencing the visceral reality of the thought experiment in a simulation, complete with emotional distress and realistic consequences.

In this episode, Chidi Anagonye, a professor of moral philosophy paralysed by indecision, finds himself placed inside the dilemma by Michael, a demon turned into an ethics student. What begins as Chidi's academic lecture on consequentialism versus deontology becomes a horrific reality as he is forced to decide in real time. When Chidi protests that the trolley problem is an abstract hypothetical, Michael responds, "Not anymore!" Chidi's trauma after several simulations shows the gap between theoretical ethics and embodied moral experience. This comedic scene may not be philosophy itself; however, it shows how cinematic thought experiments can work by engaging the audience emotionally and with a more accurate psychological experience.

This acknowledgement of the value of emotion for thought experimentation can also be furthered when we consider McClelland's response to the "problem of proper domains", which is that "the situations described by thought experiments might be beyond the proper domains of the relevant psychological capacities" (Machery 2011, 202). Machery argues that thought experiments are problematic when they describe fantastical situations, far removed from the proper domains of our cognitive capacities, and that "the reliability of a skill or capacity depends on the circumstances in which it is applied" (McClelland 2019, 108). McClelland transforms this problem into an opportunity for cinematic thought experiments. He does this by arguing that both traditional academic thought experiments and cinematic ones exist on a "spectrum of fancifulness" (Ibid., 109). Not all films are equally removed from everyday experience. Furthermore, cinematic thought experiments present scenarios within a rich, engaging narrative, not in abstract form, and this is closer to the ordinary domain of our common psychological capacities. Given that the judgements we make in our real lives are formed in similar rich perceptual environments, films may approximate this better:

> Perhaps the proper domain of our capacities for judgement is scenarios embedded in a wider reality: we never make judgements about situations that somehow exist in isolation from a wider world. The fragments of possible worlds we find in academic TEs would fall outside this domain, but filmic depictions of scenarios embedded in full narratives and rich environments might fall within it. If so, our judgements about this scenario would be more reliable than if that same scenario were sketched in an academic text (Ibid., 110).

Thus, McClelland concedes that we should evaluate if a film (or any hypothetical scenario) falls within the proper domain of our judgement on a "case-by-case basis" and not by making generalisations (Ibid), regardless of whether it is, for example, a realistic or a science fiction film.

Interestingly, Knight argues that philosophical thought experiments typically involve science fiction scenarios, while artistic narratives often do not. She cites examples like Davidson's "swampman" thought

experiment[12] as a sci-fi scenario. Then Knight notes that philosophers who defend artistic thought experiments usually appeal to realistic fiction rather than sci-fi (Knight 2011, 288). This resonates with Vidmar's discussion of how science fiction has the closest resemblance to traditional thought experiments. However, Carroll challenges Knight's distinction by pointing out that not all philosophical thought experiments involve science fiction, citing examples like Searle's Chinese room and Rawls' veil of ignorance (Carroll 2017, 282). Moreover, as discussed in our earlier discussion of Vidmar's analysis, even realistic fiction can function as thought experimentation at the thematic level. It is also not true that, at least in the case of films, philosophers are usually concerned with realistic fictions. Several sci-fi films were examined by film as philosophy advocates (*The Matrix, Blade Runner, Alien*, for instance), as already mentioned.

Whether a film presents a sci-fi or a realistic scenario should not be relevant to understanding how it can work as a philosophical thought experiment. The details presented in them and the psychological mechanism behind them are what is important to understand their potential philosophical value. It is undeniable that emotions play a fundamental role in our everyday judgments. In this aspect, too, films can have an advantage over abstract, textual thought experiments, as they engage the viewer emotionally. This point emerges in McClelland's analysis of *Ex Machina*. The film's philosophical value does not just lie in presenting arguments about artificial consciousness and creating an emotional engagement with the character, with artificial intelligence, Ava, who challenges our confidence in distinguishing between authentic and simulated personhood.

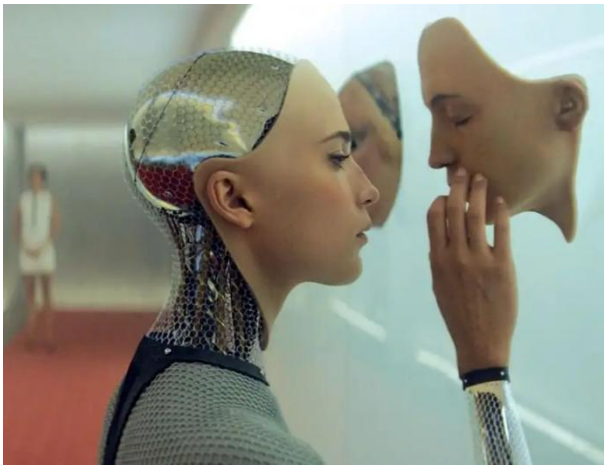


Figure 14: Ava. *New Scientist*. 2015. *Ex Machina: Quest to create an AI takes no prisoners*. Accessed April 27, 2025.

So, unlike traditional thought experiments that often simplify scenarios, films can "provoke us to reflect upon a moral issue from multiple perspectives and where the impact of actions is emotionally salient" (McClelland 2019, 104). And, as such, as Shamir claims:

[12] In this thought experiment, Donald Davidson imagines himself walking through a swamp when suddenly he is struck by lightning and killed. At the same time, by coincidence, another lightning strike rearranges the molecules of the swamp to duplicate him, creating a swampman. This duplicate looks, sounds and behaves exactly like him. However, Davidson then argues that despite being indistinguishable, swampman cannot mean or refer to anything because it lacks any history or any causal connection (Davidson 1987).

> High stakes are what create emotional engagement and empathy, which produces the 'hook' that makes the audience see, feel, and, most important, care about the characters, story, and world that revolve around a philosophical idea. If written and oral philosophy are what gave philosophy a mind, it is cinematic philosophy that gives it a heart (Shamir 2016, 127).

4.3.5 Mental modelling through film

If the philosophical thought experiment is made after watching the film, and not while during its visualisation, then we can argue that the experience of watching the film may be, at best, just a prompt, and not the philosophical thought experiment itself. It can be true that we engage with its concepts more explicitly after watching the film, which is similar to what we do with traditional thought experiments. In this latter case, we do not say that thought experiments are not philosophy just because we can and do engage with their concepts explicitly after reading and imagining them.

We recognise that the way a thought experiment is presented influences our intuitions and the way we reorganise and rethink our concepts, and philosophy is done both while imagining those scenarios and after. When a philosopher creates a thought experiment, they do so to generally "shake" our intuitions about certain concepts. It would be strange to argue that a thought experiment's only philosophical value is the discourse surrounding it, and that it does not have value by itself – a view that Norton defends and that limits the important role many thought experiments do have in science and philosophy. Drawing on Gendler's view of thought experiments to whom the details of the narrative mobilise previous knowledge, Davies maintains that the understanding achieved through cinematic thought experiments is internal to the viewing experience (Davies 2019, 144). Just as traditional thought experiments mobilise cognitive resources that we already possess but cannot articulate in propositions, films engage our understanding in ways that abstract philosophical texts cannot. A film's sensory experience allows the audience to "reflectively experience the embodied nature of the character's engagement with their lived worlds" (Ibid., 151).

Now, if we build on the mental modelling account, we can more fully understand how the experience of watching a film may constitute a philosophical thought experiment, rather than merely providing content to be extricated and explained. As analysed before, the mental modelling account describes thought experiments as dynamic simulations that engage our evolved cognitive abilities for mental representation. Unlike traditional textual thought experiments that rely on the reader's capacity to construct mental simulations, films externalise aspects of this process through their presentation. As Nersessian describes, thought experiments create "constructed situations [that] inherit empirical force by being abstracted from both our experiences and activities in, and our knowledge, conceptualizations, and assumptions of, the world" (Nersessian 1992, 297). Films extend this process by presenting richly detailed simulations that directly engage our perceptions, activating the same cognitive mechanisms involved in

processing real-world experiences. For example, when watching *Blade Runner*, we do not merely think about what it might mean to encounter a being that challenges our conception of personhood. We witness Roy Batty's emotional death, experiencing the philosophical problem through sensorial, perceptual, and emotional channels.

The distinctive characteristics of film offer these unique possibilities for enhancing mental modelling in ways that influence how thought experiments function. This enhancement comes from some main aspects of cinematic presentation: visual expression, temporal manipulation and affective engagement. Each of these aspects contributes to how specifically films can construct and present philosophical thought experiments.

The phenomenological aspects of watching a film are fundamental to understanding this. Carroll, Wartenberg and Choi note how the experience of watching films contributes to their philosophical impact. According to Wartenberg, "these films present us with deep phenomenological experiences of the nature of cinema that written philosophical texts cannot rival" (Wartenberg 2007, 132). Through techniques such as camera angles, framing and focus, film can literally embody different points of view. For example, *Rear Window* (1954) by Alfred Hitchcock confines us almost entirely to the main character, Jeffries' apartment and restricts our knowledge to what can be observed from his window. The strategy here begins with establishing the physical constraints of Jefferies' perspective. This wide shot shows his position relative to his neighbours, establishing the literal frame through which he and the viewers must gather knowledge.



Figure 15: Wide shot of Jefferies' apartment showing the rear window view of the courtyard and surrounding apartments. https://the-take.com/read/was-arear-windowa-shot-in-an-actual-apartment-courtyard-or-an-elaborate-set

Then, when Jefferies uses optical devices to extend his vision, the camera also adopts those properties.



Figure 16: Point of view binocular shot with circular matte showing what Jefferies sees when looking through binoculars

This binocular shot transforms the camera into an extension of the character's perception.

Another example is how Carpenter uses distinctive camera shots to embody several perspectives in *The Thing* (1982). The film begins by establishing visual isolation through extreme wide shots that position the human outpost as tiny and vulnerable.





Figure 17: Wide shot of the Antarctic research station from outside, emphasising its isolation in the vast, empty landscape

Carpenter also cuts between medium shots of the characters' faces when "the thing" reveals itself.







Figure 18: Sequence of medium shots of the men's shocked faces as they witness the blood test results, just before one of them (Palmer) begins to transform.

This sequence creates a visual relationship between the act of seeing and what is seen.

These techniques can be ways of activating and directing mental modelling processes. The camera guides us to different perspectives when we engage with camera perspective variations through different angles, distances, and movements.

Films are adequate to activate and reconfigure these cognitive structures. While written thought experiments have to describe the events, films can just show them, and thus directly prompt our "stored knowledge" in an immediate way.

This also fits with Gendler's constructivist account. Gendler argues that "contemplation of an imaginary scenario may lead us to new beliefs about contingent features of the natural world that are produced not inferentially, but quasi-observationally" (Gendler 2004, 1154). Film extends this quasi-observational process, presenting scenarios that viewers experience as if they are witnessing them directly. As we experience a film, we construct, test and revise mental models guided by how the film presents its content. These models integrate explicit content with tacit knowledge, emotional responses and embodied understanding. When considering the mental modelling account of thought experiments, we can see film's ability to manipulate perception, time and emotional engagement to create an especially interesting environment for philosophical thought experimentation.

4.4 An integrated account of cinematic-philosophical thought experiments

We are now able to articulate a revised bold thesis about how films can philosophise through thought experimentation. This synthesis enables us to identify some necessary conditions for cinematic thought experiments to be philosophical, while acknowledging their unique advantages and the scepticism they face.

For films to function as philosophical thought experiments, they must employ the methodological structure of thought experimentation, address philosophical subjects, and use cinematic techniques responsible for the philosophy being done. Unlike Mulhall's stronger claim that films can engage in serious and systematic philosophical thinking in "exactly the same ways" as traditional philosophical texts, our position acknowledges the medium-specific nature of cinematic philosophising while maintaining that films can philosophise. When a film like *The Matrix* presents the deception hypothesis, it reconfigures it in a way that, as Wartenberg argues, places viewers in a sceptical position that text alone cannot.

The advantages of cinematic thought experiments are particularly evident when we consider some critiques of traditional thought experiments for their failure to specify adequate background conditions. Through their audiovisual nature, films provide detailed context and worlds that can address this. For example, when *Eternal Sunshine of the Spotless Mind* explores the role of painful memories in personal identity and interpersonal relationships, it does so against a fully realised background that enables the audience to consider implications that might remain unclear and vague in a purely textual thought experiment. As McClelland notes, "film presents us with scenarios we can see and hear so might better fall within the proper domain of our judgment" (McClelland 2019, 110). This experiential dimension transforms the traditional thought experiment in ways that Shamir emphasises when he argues that film does not ask its audience to merely imagine an idea but lets them "see and experience the idea in action" (Shamir 2016, 112).

This experiential quality addresses the sceptical positions against thought experiments that we examined earlier. The gap between hypothetical and real cases can be narrowed by film, while the psychological limitations of traditional thought experiments might be, at least, partially overcome through film's engagement with several cognitive faculties. When films activate emotional and intellectual responses simultaneously, they can mobilise our mental models, the "quasi-observational" processes that reconfigure our conceptual understanding.

We must also confront the dilemma of paraphrase: if a film's philosophical content can be adequately paraphrased in traditional philosophical texts, then the film itself seems superfluous; if it cannot be paraphrased, then articulating the film's philosophical contribution becomes problematic. Our position resolves this by recognising that paraphrasability does not necessarily entail a loss in philosophical content. While a film's philosophical insights can be articulated in prose, this does not make the film philosophically superfluous. Just as a traditional thought experiment like Thomson's violinist might be summarised without capturing its full persuasive nature, a cinematic thought experiment's philosophical content might be paraphrased without reproducing its impact. In traditional thought experiments, a paraphrase can have an impact on its content and the intuition they "pump". In the case of films, this impact is significantly greater for all the reasons previously explored.

This position avoids both the extreme claims that no film can philosophise, or that they philosophise in the same ways as traditional texts. It acknowledges the contributions of film while maintaining connections to the broader philosophical tradition. Films do not need to include explicit argumentation to philosophise. Indeed, as my analysis of thought experiments has shown, the epistemic value of thought experiments often derives from their ability to activate tacit knowledge and reorganise conceptual understanding, not from their ability to be reconstructed as explicit arguments.

The rich detail and emotional engagement that cinematic thought experiments offer are not incidental to their philosophical function but constitute advantages over traditional ones. As Shamir argues, "once the cinematic thought experiment has been manifested and actualized onscreen, it is transformed into a new and different possibility" (Ibid., 82). He argues that Harold Ramis' *Groundhog Day* (1993) does not illustrate Nietzsche's eternal recurrence but enables viewers to engage with the concept experientially: "the essence of the philosophical idea that is evoked by the film *Groundhog Day* can be understood without the help of a theoretician connecting it to its 'proper' philosophical text" (Ibid., 110).

Our revised idea, thus, maintains that films can function as autonomous philosophical works through thought experimentation. They are autonomous because they engage with philosophical problems through their unique capacities, often offering different perspectives on one philosophical subject. Thus, we adopt what McClelland calls the "Best Tool Principle", where rather than making general claims about film's philosophical potential, we recognise that "for some jobs, academic texts will be the best tool, giving us a TE with the right kind of details to improve our philosophical knowledge. But in other cases, film will be better suited to giving us the kinds of detail that need to be considered to achieve the relevant philosophical understanding" (McClelland 2019, 102).

This chapter has established the relationship between artistic works, particularly films, and philosophical thought experiments. We began by examining how artistic works can function as philosophical thought experiments through various approaches, drawing on perspectives from Vidmar, Elgin, John, Camp and Davies. These accounts demonstrate how art can engage in philosophy.

I then addressed the rationality objection and Livingston's bold thesis, which presents challenges to film as philosophy. The bold thesis requires films to make philosophical contributions through means exclusive to cinema (means conditions), and to offer historically innovative and independent philosophical insights (results condition). This led to the dilemma of paraphrase: either a film's philosophical content can be paraphrased, making it no longer exclusively cinematic, or it cannot be paraphrased, casting doubt on its philosophical nature.

I resolved these challenges by reconsidering the notion of medium specificity and arguing that films do not need to employ unique techniques but rather characteristically cinematic ones to philosophise. I

demonstrated that paraphrasability does not undermine a film's philosophical value, just as our ability to paraphrase traditional philosophical texts does not diminish their originality.

The chapter then examined how films can serve as philosophical thought experiments by addressing sceptical positions about thought experiments in general. We demonstrated that cinematic thought experiments have advantages over traditional textual ones. Their rich detail helps overcome the problem of inadequate background conditions raised by Wilkes and others. Their emotional engagement does not operate as a distraction from philosophical understanding. Rather, it provides a valuable dimension of such understanding. This view aligns with what Nussbaum and Gendler have argued about the role of emotion in cognition.

I showed that films may better engage our psychological capacities in their proper domains by presenting scenarios in rich perceptual environments similar to how we encounter real-world situations.

Through mental modelling accounts drawn from Nersessian and Miščević, we showed how films can activate tacit knowledge and reorganise conceptual understanding. The distinctive characteristics of film, such as visual expression, temporal manipulation, and affective engagement, enhance these mental modelling processes in ways that textual thought experiments cannot achieve.

Finally, I established a revised bold thesis that acknowledges the unique contributions of cinematic thought experiments while maintaining connections to the broader philosophical tradition. This approach recognises that different philosophical problems might be better addressed through different media, following McClelland's "Best Tool Principle", opening the possibility for future studies. Rather than claiming films philosophise exactly like traditional texts, we argued that they can make distinctive and valuable philosophical contributions through their characteristic means of expression, often constituting the most effective medium for exploring philosophical concepts.

Before demonstrating how films function as philosophical thought experiments through detailed case studies, I will address a final challenge posed by the imposition objection.

5. The imposition objection and authorial intentions

We now encounter a significant challenge to film as philosophy: the "imposition objection". Its idea is that philosophers are those who put or impose their philosophical interpretation and ideas onto the film: "The philosophizing is being done by a philosopher using a film, but not by the film itself" (Wartenberg 2007, 25). For Mulhall, this is a strange accusation since "films – like novels, plays and paintings – are the products of intentional human practical activity, can have representational content and can take pretty much anything as their subject matter" (Mulhall 2008, 131).

When we consider imposition and interpretation, broader questions emerge about how meaning is constructed in communication. To clarify this, we might briefly examine Paul Grice's theory of meaning.

Firstly, Grice distinguishes between two types of meaning: natural and non-natural. Natural meaning is fact-based and entails truth. For instance, when we say, "those spots mean measles" (Grice 1957, 177), we are indicating a factual relationship that cannot be false. So, Grice characterises natural meaning as that which arises directly from natural signs or phenomena. A situation where "smoke means fire" is a natural sign where the meaning is derived from an observable, causal relationship without any necessary human intention. In natural meaning, the connection between the sign and what it signifies is automatic and does not involve any individual's intention to convey a message.

Nonnatural meaning, in contrast, is intention-based and depends on human communication. When we say "three rings on the bus bell mean the bus is full" (Ibid), we are describing a conventional signal that could be false or used incorrectly. So, nonnatural meaning refers to meaning in contexts where something is meant in a communicative sense, such as when someone uses language to convey thoughts or feelings. If someone says, "That cloud means rain," the speaker intends to convey to the audience that rain is expected. Here, the meaning is tied to the speaker's intention and the audience's ability to recognise that intention. Grice emphasises that for something to "mean" in this nonnatural sense, there must be an intention in the utterer to produce some effect or belief in an audience, and this intention must be recognised as such by the audience (Ibid., 378-79).

This distinction is important for understanding artistic meaning. Like linguistic communication, artistic works operate in the realm of nonnatural meaning. They depend on human intentions and conventions rather than natural causation. Just as we understand words and

sentences through their intended communicative function, we understand artworks through their role as intentional acts of communication. So, this theory depends on some important elements: the speaker has an intention to produce a belief or response in the audience; they also intend for the audience to recognise this intention; and the audience's recognition of this intention is part of understanding the message. This helps explain how artistic works, including films, can create meaning. When a filmmaker makes specific creative choices, they are engaging in an act of communication that requires an intention to convey specific ideas, an expectation that viewers will recognise these intentions, and the viewers' engagement with these intentions as part of their interpretation.

While Grice's theory provides an understanding of how meaning operates in communication, artistic works present some challenges that extend beyond this basic of intentional meaning-making. As Vitor Moura's analysis demonstrates, artistic communication, particularly in film, involves layers of meaning-making that build upon Grice's notion of non-natural meaning. There are several differences between artistic communication and ordinary language that should be noticed. The relationship between author and audience lacks the immediacy of conversation. Moura references Proust, who distinguished between "conversation" and "reading" (Moura 2021, 16) and argues that the reader may become involved with someone else's "communicative vehicle" more intensely than in actual conversation. Different from direct conversation, artists and audiences do not share the same timeframe and cannot engage in back-and-forth exchanges. These differences create an asymmetry in the relationship.

Moura argues that watching films activates cognitive tools we typically use when interacting with people in our daily lives, though. When viewing a film like Hitchcock's *Psycho* (1960), we spontaneously reconstruct the "intentional agenda" (Ibid., 8) that framed its creation. In his analysis of *Psycho*, Moura examines how Hitchcock's camera movements guide viewer interpretation. After Marion steals money and changes clothes in her hotel room, the camera points toward the bed and zooms into an envelope of cash, then into a suitcase filled with clothes.

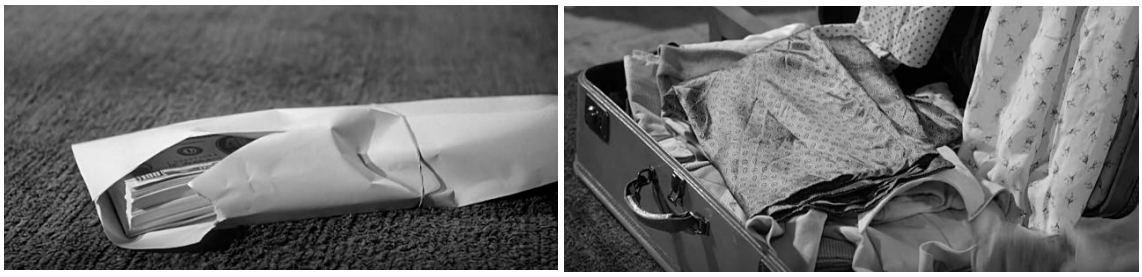


Figure 19: Camera movement highlighting the envelope of stolen money and a packed suitcase.

According to Moura, this camera movement serves as a cue that prompts viewers to interpret that Marion is stealing the money and fleeing town, rather than simply packing for a weekend (Ibid., 10). He also analyses the sequence following the shower scene, where Hitchcock shows the bathtub draining, Marion's dead eye, and a folded newspaper containing Marion's stolen money.



Figure 20: The shower scene sequence.

These visual elements contain motifs that resist straightforward verbalisation. The sequence's visual density creates a communicative experience that transcends simple verbalisation (Ibid., 14-15). Through these techniques, Hitchcock exemplifies what Moura calls the "Spectral Author", a presence that emerges from the work itself yet retains a connection to the real director, fostering a sense of intimate communication between creator and audience. This spectral presence allows viewers to feel they are engaged in a form of communication that paradoxically may be deeper precisely because it lacks conversation's immediate demands (Ibid., 18-19).

The separation, complex symbolic means, and multiple levels of meaning show how artistic communication has evolved methods of conveying meaning despite its distance from ordinary conversation. This suggests that understanding artistic meaning requires attention to not only the basic principles of communicative intention but also the specific ways that art, and film specifically, has developed to transcend the limitations of indirect communication.

An important aspect of Grice's theory is that conventional meanings develop over time from the speaker's intentions. In art, this helps explain how artistic conventions develop from creative practices; these conventions shape but do not fully determine meaning, and new meanings can emerge through innovative uses of conventions. This understanding helps us navigate between two extremes in interpretation: a pure intentionalism that reduces meaning to biographical facts about the artist, or pure conventionalism that ignores the role of creative intentions.

With this understanding of artistic meaning in place, we can better analyse the imposition objection. The question is not simply whether philosophers impose their interpretations on films, but instead how we can interpret films as communicative acts while respecting both their intentional and conventional aspects. So, the imposition objection raises an important question about film interpretation: how can we distinguish between legitimate interpretation and philosophical impositions?

When we interpret a film philosophically, we are not just identifying content but making claims about communicative intentions. The worry about imposition can be reframed as a concern about misidentifying these intentions. Are we recognising genuine communicative intentions in the film, or are we attributing intentions that were never there? This shows why the imposition objection is important. It is not simply about whether films can contain philosophical content. Instead, it is about how we identify and interpret it. When we claim a film is "doing philosophy," we are claiming that there are intentional communicative acts within the film; these acts are meant to convey philosophical content; and the audience is meant to recognise and engage with this philosophical meaning.

Before we try understanding authorial intentions, though, we should address the fact that not all philosophers agree that we need to know the author's intentions in order to understand meaning. The intentional fallacy, first proposed by W.K. Wimsatt and Monroe Beardsley, is one of the most significant challenges to intentionalist approaches to interpretation. The fallacy refers to the mistake of using the author's intentions as the standard for judging or interpreting a work of art. As Wimsatt and Beardsley argue, "the design or intention of the author is neither available nor

desirable as a standard for judging the success of a work of literary art" (Wimsatt and Beardsley 1946, 468).

5.1 Wartenberg's response: creator-oriented interpretation

Wartenberg firstly responds to the imposition objection by saying that this kind of general and *a priori* objection should not be made; instead, he argues that we should ask the appropriateness of each interpretation for a specific film. For the philosopher, it is possible that some specific interpretations impose ideas on the film that are not indeed present in it, but that does not mean that this happens in all cases of interpretations of film as doing philosophy.

Indeed, Wartenberg criticises, for example, Noël Carroll's interpretation of *The Third Man* for considering that he imposes on the film an interpretation that the film would not allow. Carroll sees this film as a counterexample to the maxim "When loyalty to a friend conflicts with loyalty to a cause, one ought to choose in favour of the friend" (Carroll 2002, 10). Wartenberg, however, thinks that this film is, on the one hand, a thought experiment that seeks to develop the problematic of the "vicissitudes of moral intelligence" and, on the other hand, an illustration of the Aristotelian approach to friendship (Wartenberg 2007, 94). So, for the latter, this film must be interpreted as making a stronger philosophical contribution than Carroll thinks – although it can be said that both philosophers agree, at least in general, that the central theme is friendship.

Thus, for Wartenberg, the imposition objection may be useful, not as a general and *a priori* objection that precludes film as philosophy, but rather to construct a principle: "When making an interpretation of a film, take care not to impose an interpretation on a film that is inappropriate" (Ibid., 26).

By an "inappropriate" interpretation, Wartenberg means an interpretation that imputes in the work a meaning that its creator could not have intended. As such, the solution to avoid making inappropriate interpretations that Wartenberg suggests is a kind of "creator-oriented" interpretation, instead of a kind of "audience-oriented" interpretation. Audience-oriented interpretations are freer, unrestricted by what the author might have actually meant. Such types of interpretations can use any kind of artwork for the intended purpose, such as the exploration of some type of philosophical question. This distinction between creator and audience-oriented interpretations inevitably recalls Roland Barthes' "Death of the Author" (1967). Barthes argued that treating the author as the source of meaning was to "impose upon that text a stop clause, to

furnish it with a final signification" (Barthes 1967). While this view might seem to support unlimited audience-oriented interpretations, it does not require us to abandon all consideration of authorial intent.

A more moderate approach found in Wartenberg's creator-oriented interpretations acknowledges that while meanings can extend beyond an author's explicit intentions, not all interpretations are equally valid or defensible. The key is not to treat the author as an infallible authority, but to use their potential intentions as a reasonable constraint on interpretation, preventing readings that would have been impossible in the work's historical and creative context. We find that this approach allows us to be rigorous in our interpretation. However, it also helps us avoid two extremes: giving complete control to the author's explicit intentions or allowing interpretation to become chaotic.

Creator-Oriented interpretations, though, have the main limit of not affirming something that the author could not have intended to affirm. Wartenberg gives us the following example:

> Seeing an artwork from the seventeenth century addressing Einstein's theory of relativity makes no sense as a creator-oriented interpretation, even if it can be a profound audience-oriented one, for the obvious reason that no one living in the seventeenth century could have had any knowledge of Einstein's theory. (Ibid., 26)

However, the philosopher notes that the author of a work does not have to be aware of everything that is involved in a thesis, which, for example, she seeks to illustrate in her film. Thus, the author of *The Third Man* would not have to be familiar with Aristotle's work in order to be able to present a thesis on friendship and the vicissitudes of moral intelligence: "because filmmakers exist in a broader cultural and intellectual context, they are aware of some philosophical issues prominent in their society, which animate both philosophical texts and its more general culture" (Wartenberg 2007, 139).

This creator-oriented approach provides a method for interpretation that respects the filmmaker's context and potential intentions, prevents wild, unsupported interpretations, and establishes a way of assessing philosophical content in films.

5.2 Livingston's critique of Wartenberg

Paisley Livingston critiques Wartenberg's perspective. Regarding Wartenberg's idea that when the expression "films philosophise" is used, it means that filmmakers philosophise through their films. His response to the imposition objection, Livingston states that "difficult issues are raised by reference to film authorship, as well as by the postulated link between authorial intentions and cinematic philosophizing." (Livingston 2008, 593). One such problem is the fact that films often have multiple authors, a problem that Wartenberg referred to as the "no-author objection" (Wartenberg 2016). Films usually have several authors and an entire crew. The final product of a film, particularly a typical narrative film, is the result of the efforts of one or more screenwriters, producers, designers, editors, actors, composers, and, of course, directors, among others. Therefore, we "cannot immediately identify the director as the one individual who is solely responsible for the final product" (Wartenberg 2016, 176).

A possible response offered by Wartenberg is to argue that, although they are the result of a collective work, films can have a united intention. Wartenberg also points out that there are traditional ways of philosophising that are not subsumed to a single author. For example, he suggests verbal discussions such as the Socratic method that require at least two participants that did not necessarily share the same point of view. We might also add that other traditional philosophical ways, as writing books or articles, may be the result of the joint efforts of various authors or contributors with a common intention to convey specific philosophical ideas; the same can be said of conferences, workshops, lectures, and so on.

Livingston himself resolved this problem by arguing that authorship requires the author(s) to have "sufficient control over the production of a work" (Livingston 2009, 70). That is, if the creator(s) have sufficient control over the final results, it is possible to think that they have an intention that, if achieved, can be interpreted. This is one of the main reasons why Livingston argues that Bergman's work can be considered as philosophising: as is known, Bergman had good control over the final product of his films.

Nonetheless, Livingston questions the response that Wartenberg gives to the imposition objection based on the creator-oriented interpretation that relies on investigating what the creator could have intended (or avoided what he could not have intended). According to Livingston, the very expression "could have intended" is too ambiguous (Livingston 2008, 594). For the philosopher, Wartenberg's idea that it is not necessary for the author to know the philosophical texts related to the subject she addresses in her film, for example, arguing that it would suffice to

know the cultural context in which the author is inserted, is not enough to make a correct interpretation of what the author could intend.

To demonstrate his point, Paisley Livingston refers to Wartenberg's interpretation of *Eternal Sunshine of the Spotless Mind* (Michel Gondry, 2004) as an example. The latter had argued that this film is a specific type of thought experiment (counter-example), which serves to counter-exemplify utilitarianism. Livingston claims that there is no evidence of this being the author's intention: Wartenberg "has no specific evidence indicating that Gondry, the scriptwriter, or Gondry's other collaborators actually read or thought about [the book] *Utilitarianism* or related works in moral philosophy" (Ibid., 594).

The main problem that Livingston identifies in Wartenberg's suggestion is that it is too loose or broad, to the point of allowing too many contradictory interpretations: "Are we to conclude that the philosophizing 'done by the film-maker' in the film embraces a multiplicity of incompatible assertions?" (Ibid., 595).

5.3 Types of intentionalism

The type of solution defended by Wartenberg, which uses the author's potential or hypothetical intentions, is, Livingston claims, a specific type of intentionalism, that is, a thesis that seeks to determine the intentions of the author. Livingston defines the author's intentions as "executive attitudes towards a plan regarding the intending party's own future actions," which implies that the author has a specific plan upon which he intends to act and hence "in the absence of a plan, there can be no intending" (Livingston 2009, 85).

5.3.1 Conditional intentionalism

The kind of intentionalism that Livingston thinks that Wartenberg defends is the *conditional intentionalism* that "invites the interpreter to describe the meanings that the author of the text (or the director of the film) *could have* intended, as opposed to those that this person actually intended" (Ibid., 87). Conditional intentionalism offers us a different way to approach interpretation. When we use this approach, we ask what meanings the author of a text (or the director of a film) could have intended, rather than what they actually intended (Ibid). This approach focuses on possible rather than actual intentions.

Livingston points out issues with this type of intentionalism that, he argues, are evident in Wartenberg's thesis. According to Livingston, this kind of conditional intentionalism cannot determine what counts as valid evidence for possible intentions. Wartenberg's approach suggests that it is sufficient for specific ideas to be in "general circulation within a culture" (Ibid., 90) to be considered as possibly intended by the creator. However, as Livingston argues, this criterion may be too permissive, even warranting incompatible readings without sufficiently determining a coherent result. For example, according to Livingston, because of different versions of utilitarianism circulating in culture, *Eternal Sunshine of the Spotless Mind* may warrant contradictory interpretations under conditional intentionalism (Ibid).

5.3.2 Hypothetical intentionalism

Hypothetical intentionalism seeks to investigate the social context in which the author inserts themselves, considering it unnecessary to resort to sources that can give clues on the intentions of the author, such as letters or interviews (Ibid., 92). It seems that Wartenberg's solution would also fall into this category.

This perspective can be linked to Jerrold Levinson's. Hypothetical intentionalism shows us that interpretation involves uncovering intentions, but not in the way we might expect. Rather than seeking the artist's true intentions, we focus on developing the most credible theories about what those intentions might have been. We build these theories from the artwork itself and some external context.

According to the hypothetical intentionalist perspective, the meaning of a text is determined by what an ideal reader would hypothesise as the intended meaning. This ideal reader must be fully informed about several elements: the cultural background of the text, the author's complete body of work, all publicly available information about both the text and its author, and, of course, the text itself. A crucial aspect of this approach is its strict limitation on permissible evidence: only information that would be publicly available to the intended, appropriate reader is allowed, while private sources such as interviews with the artist or personal papers are explicitly excluded from consideration.

Levinson distinguishes between what an author actually intended and what a reasonable reader would justifiably infer the author intended. As he explains, hypothetical intentionalism "centers on the idea that a literary work should be seen as an utterance, one produced in a public context by a historically and culturally situated author, and that the central meaning of such a work

is thus a form of utterance meaning, as opposed to either textual meaning or utterer meaning" (Levinson 2010, 139). He identifies distinct models of literary meaning. Textual meaning refers to "the meaning of the brute text as a string of words in a language"; utterer meaning is the meaning that the "utterer, speaker, or author had in mind and intended to get across" (Ibid); the utterance meaning, in contrast, is what a reader would most reasonably take the author to be trying to convey through their text within its context (Levinson 2006, 295). So, for Levinson, utterance meaning is the "key" to literary interpretations. As he states, "it is a best hypothesis about authorial intent, and not authorial intent *per se*, that is constitutive or determinative of central literary meaning", emphasising that "the core meaning of a literary work is given by the best hypothesis, from the position of an appropriately informed, sympathetic, and discriminating reader, of authorial intent to convey such and such to an audience through the text in question" (Levinson 2010, 139). This would allow us to interpret literature or film as communication while avoiding the problem of accessing actual mental states. The focus shifts from recovering what the author meant to constructing the most plausible account of what the author intended.

This approach faced criticism, particularly from actual intentionalists like Livingston or Carroll. Carroll argues that if new evidence, such as an artist's notebook, reveals an author's actual intent, this should take precedence over any previously well-warranted hypothesis. He also points out that readers typically behave as if they are identifying actual intentions rather than merely making hypotheses.

5.3.3 Actual intentionalism and partial intentionalism

Noël Carroll defends what he terms "modest actual intentionalism", which he distinguishes from extreme versions of the theory. According to him, an extreme actual intentionalism maintains "that the meaning of an artwork is fully determined by the actual intentions of the artist" (Carroll 2000, 76). This leads to "Humpty-Dumpty-ism", which is "the idea that an author could make a work mean anything simply because they will it so" (Ibid). By contrast, a moderate actual intentionalism only claims that "the artist's actual intentions are relevant to interpretation" and "constrain our interpretations of artworks," rather than fully determining meaning. Specifically, with reference to literary texts, "the modest actual intentionalist argues that the correct interpretation of a text is the meaning of the text that is compatible with the author's actual intention" (Ibid). Thus, this thesis avoids Humpty-Dumpty-ism since "the intentions of authors that the modest actual intentionalist takes seriously are only those intentions of the author that the linguistic/literary unit can support

(given the conventions of language and literature" (Ibid). His thesis rests on the claim that artistic interpretation should be continuous with ordinary interpretation.

Carroll also addresses common objections to actual intentionalism. Against the charge that intentionalism diverts attention from the text, he responds that the best evidence for what an artist intend is still the artwork itself, and that the attention to the text itself is the pursuit of actual intention. Thus it is incorrect to think that paying close attention to the artwork and searching for actual intentions are opposite (Ibid., 77). Regarding the fallacy of paraphrase, Carroll clarifies that "the aim of the modest actual intentionalist is not primarily to return home with a paraphrase of the author's intention [...], but to use the author's intended meaning as a resource for engaging the text" (Ibid).

Robert Stecker also defends what he calls the "unified view" of moderate actual intentionalism, arguing that "text-based works" have the same sort of meaning as other linguistic utterances and that reference to the actual intentions of a work's creator plays an essential role in specifying what that meaning is. Stecker defines a successful intention:

> an utterer *successfully* intends a meaning X just in case the utterer intends X, the utterer intends that the audience will grasp this in virtue of the conventional meaning of the utterer's words or an extension of the meaning permitted by those conventions, and the first intention is graspable in virtue of those conventions or permitted extensions of them (uptake condition)" (Stecker 2006, 429)

He also argues that "utterance meaning is not always intended meaning" (Ibid) allowing cases where communication fails.

Stecker also addresses the thesis that we should eliminate intentions from meaning, arguing that they cannot be eliminated because "we will not be able to recognize irony, allusion, parallelism, imagery, or symbolism without being directly involved in figuring out a communicative intention" (Ibid., 431) Stecker also responds to claims that intentions are unknowable by pointing out that "there are times when it is easy to know what others intend". He argues that if we say "Do you want to go for a walk?" after dinner, "you know that I intend to ask you whether you want to go for a walk" (Ibid 435).

Livingston defends a partial intentionalism, which he claims to be a form of actual intentionalism. This approach "holds that intentions determine some, but not all, of the semantic properties of at least some works of art" (Livingston 2009, 93). He introduces the concept of

"meshing" as the idea that for an author's intention to be successful, it must sufficiently mesh with what is presented in the work. Livingston offers a method for evaluating interpretations. He suggests that interpreters should investigate both "internal" evidence (the characteristics presented in the work) and "external" evidence (such as context, notes, or interviews). This process should be one of reciprocal adjustment, "leading ideally to a reflexive equilibrium based upon an integration of the totality of relevant evidence" (Ibid., 108). He also acknowledges the potential limitations of this approach. Since interpretations involve empirical and historical investigation through external sources, they may fail. There might be cases where external evidence is missing, and "questions about both intentions and their success cannot be answered" (Ibid., 121). In such situations, Livingston accepts that the interpreter may need to hypothesise about implicit meanings in the work.

5.4 Livingston's interpretation of Bergman's work

After arguing that we need to understand authorial intentions to comprehend a film's potential philosophical meaning, Livingston tried to understand Bergman's intentions. Towards this, he traced what he argued to be the best path to do so: he analysed both internal and external evidence and tried to see if everything meshes. That was when he identified a philosophical source or inspiration in Bergman.

Livingston affirms that "although it is not widely known, Bergman read and was strongly influenced by the Finnish philosopher and psychologist Eino Kaila" (Ibid., 126). He also quotes Bergman, who actually stated in an interview about *Wild Strawberries* [*Smultronstället*] (1957):

> Philosophically, there is a book which was a tremendous experience for me: Eino Kaila's *Psychology of the Personality*.[13] His thesis that man lives strictly according to his needs – negative and positive – was shattering to me, but terribly true. And I built on this ground (quoted in Livingston 2009, 126)

So, Livingston makes a summary of Kaila's positions that he identifies in Bergman's films. In what follows, there are some of these main positions.

Firstly, there is the idea that there is only one effective psychic force that is in Finnish called "*tarve*", translated as need or desire. Aside from that, there is a conflict between 1) *basic appetites*

[13] In the original Finnish text, *Persoonallisuus* (1934).

and 2) s*piritual* or *high spiritual desires*, such as the desire for religious purity or salvation, and living systems have a tendency to move (or try to move) from a state of disequilibrium to a preferred state of equilibrium (Ibid., 129-30).

We are also faced with what Kaila calls "inauthenticity", that is, the rejection of rationalist conceptions of the mind, focusing on the decisive influence of motivational forces on cognition and belief. For example, to avoid feelings of humiliation or low self-esteem, humans shift their beliefs about states of affairs irrationally. When the fox tries to reach the grapes and does not get there, then she irrationally shifts her belief to assume that the grapes are sour anyway (Ibid., 130). This does not mean that rational desires do not exist, but he emphasises the tendency for humans to satisfy irrational desires.

Furthermore, Kaila speaks of "displacements of motivation", referring to what happens when a desire cannot be readily acted on or satisfied. On this occasion, it may be replaced by a desire for some surrogate object. One example is the sadistic punishment of children: such a desire is inauthentic since its basis is another aggressive desire of the adult (Ibid., 131).

Another important aspect is Kaila's contention that "one of the best ways to gain insight into the working of normal human personalities is to study cases where the psychic system malfunctions and breaks down" (Ibid., 136). In this regard, Kaila explores the situations, such as the schizophrenia case, where distinctions between, for example, dream and reality collapse, a collapse "characteristic of magical thinking" (Ibid., 137).

Finally, Kaila believes that self-awareness is inauthentic and irrational, and sometimes attempts at self-reflection turn into wishful thinking. One possible way of becoming somewhat lucid, however, is to avoid social norms and ideals and then reflect (Ibid., 168).

Now, Livingston argues that "some of Kaila's main themes correspond quite well to important aspects of Bergman's own experience, such as childhood humiliations and ritualized punishments, as well as his powerful ambitions, sexual desires, and notorious outbursts of anger" (Ibid., 134). Besides this, he adds that "Bergman's artistic *oeuvre* manifests a lifelong interest in, and condemnation of, the spontaneous, ritual, and artistic forms of scapegoating and humiliation" (Ibid.).

Livingston analyses some of Bergman's films and how their philosophical elements are inspired by Kaila, and finds the exploration of the topic of magical thinking. First, *Through a Glass Darkly* [*Såsom i en spegel*] (1961) revolves around a schizophrenic character, Harriet Andersson,

who suffers from horrible hallucinations, a topic that is crucial for Kaila's treatment of functional and dysfunctional psychic systems and the collapse between dream and reality.

In *The* Magician [*Ansiktet*] (1958), a character (Granny Vogler) casts a spell over servants and children. Moreover, in the film, the magician Albert Emanuel Vogler tricks the rationalist Minister of Health, Dr Vergerus, which illustrates Kaila's assertion that "the animistic tendency to mistake purely natural events for the working of spirits can manifest itself when the rational, educated person experiences a crisis of sufficient magnitude" (Ibid., 137).

In *Persona* (1966), we also find an instance of magical thinking, as well as a kind of displacement of motivation. Kaila thinks that a case of magical thinking would be, for example, when someone tears a picture in two, venting their emotions about a person into a piece of paper. This is exactly what happens when the actress, Elisabeth Vogler, anxiously tears the picture of her son in two, an irrational action that displaces the character's angst about her child onto a mere object. This displacement of motivation, the attack against a surrogate object or the scapegoating seems to be what happens in *From the Life of the* Marionettes [*Aus dem Leben der Marionetten*] (1980) too, where Peter Egerman fantasises about killing his wife, Katarina, and ends up killing a prostitute, Katarina Krafft. We also see scapegoating in *The Seventh Seal* [*Det sjunde inseglet*] (1957) as we watch a child being ritually sacrificed.

In *The Silence* [*Tystnaden*] (1963), we find a conflict between what can be seen as "high" and "low" desires, where "Esther (...) says she is 'humiliated' by her sister Anna's (...) promiscuity, whereas Anna is angered by her sister's high-minded and accusatory attitude about her sexual activities" (Ibid., 140).[14]

Livingston states the following:

> While I do not propose to try to reduce Ingmar Bergman's fictions to simple illustrations of Kaila's ideas, I think it important to recognize the extent to which the characterizations and rhetorical design of Bergman's cinematic works mesh powerfully with some of Kaila's central psychological premises." (Ibid., 127)

This is the almost classic avoidance of saying that a film merely illustrates a philosophical thesis that was analysed before, and this seems also to be especially worrying if Livingston wants to maintain a bold thesis in which originality and independence are requirements.

[14] More examples with more detailed analyses and extensive treatment of Kaila's philosophy can be found in chapter 5 of Livingston's *Cinema, Philosophy, Bergman* (2009).

Livingston only seems to clarify how Bergman goes beyond Kaila in his films when accounting for Bergman's views on morality. As he says, in Bergman, forms of evil are not disqualified as inauthentic or irrational, as Kaila would think. Instead, Bergman states that "there exists an evil that cannot be explained – a virulent, terrifying evil – and humans are the only animals to possess it" (quoted in Livingston 2009, 187). Still, this gives the overall impression that Livingston attaches Bergman's work too closely to Kaila, not being very open to other possible interpretations anyhow.

Here is an example of what Livingston thinks about other types of interpretation:

> Ingmar Bergman is notorious for having been a 'gloomy' film-maker and has often been vaguely associated with existentialism. The many published interpretations of his films draw on a number of strikingly different theoretical backgrounds, including the work of the Danish philosopher Kierkegaard, Sartre, Camus, Freudian and Jungian psychoanalysis, Wittgenstein, and Lutheran theology. (...) While Bergman's films can, of course, be interpreted by applying notions drawn from these and various others jointly incompatible theoretical sources, it has never been established that Bergman's artistic projects were actually inspired of informed by them. (Ibid., 125)

In parenthesis Livingston says that "this is not meant to rule out, by the way, the unintended meaning and significance of the works" (Ibid., 128), and continues to say that "although there are some Freudian elements in Kaila's work, he was sharply critical of various psychoanalytic tenets and classified Freud's mode of thought as an outdated 'mechanistic associationist psychology'" (Ibid., 129). Kaila rejects Freud's emphasis on sexuality as well, along with Nietzsche's emphasis on will to power, rejecting the issue of dreams, castration anxiety, penis envy, Electra complexes, and so on (Ibid., 132).

Livingston maintains that the main reason why there are not many scholars bringing Kaila to interpretations is because his work, originally written in Finnish, was never translated into the main languages (like English, French or Spanish), which seems to be an implication that other interpretations are wrongly done due to a lack of that knowledge.

Essentially, Livingston's somewhat controversial position seems to be that: a) there is no proof that Bergman had existentialist influences (for example, he says that there are no references to Kierkegaard for what he means by "angst"; and b) Bergman had no interest in defending psychoanalysis, since he was influenced by Kaila who rejected that tradition.

The above thesis seems to exclude from the possible realm of good interpretations any existentialist or psychoanalytic accounts, for the reasons we saw above. This seems strange to many scholars, for example, Dan Williams, who states:

> Livingston draws many interesting parallels between Kaila's ideas and Bergman's work, including the psychological consequences of corporal punishment on a child (...) but it can be noted at the outset that some of these ideas overlap with those of other thinkers. Of particular interest here, is the notion that Bergman shares Kaila's interests in magical thinking and the way in which individuals are driven by irrational motivations. These are both explicit themes in the early work of Jean-Paul Sartre. For example, Sartre conveys the obsessive dynamic of emotion with reference to "magical thinking" (Williams 2015, 165)

This seems to indicate that there is an actual overlap of interests between Kaila and Sartre (and, of course, Bergman).

The major problem seems to be, as Murray Smith notes,[15] that Livingston isolates Kaila's work from its broader context, which was, indeed, imbued with existentialism.

The same could be said about Livingston's rejection of psychoanalytical interpretations, because Kaila mostly rejects psychoanalysis. And, since Kaila influenced Bergman, he rejected it, too, it seems. But there are important notes to mark here: influence does not mean acceptance. It could be the case that Kaila was influenced by Freud precisely because he closely studied it. To be influenced by an idea does not mean endorsing it in every way, whether intended or unintended, conscious or unconscious. Besides, even if Bergman did reject Freud's psychoanalysis, why wouldn't he try out a psychoanalytical content, even if not to show that it is not a reasonable account of the mind, for example, portraying bad examples of psychoanalysts and psychoanalytic explanations?

It seems that Livingston's partial intentionalism has the risk of failing to acknowledge that there may be actual existentialism or psychoanalysis in Bergman, because he is too attached to Bergman's explicit endorsements, Kaila's clear positions, and Livingston seems to be ignoring other important factors. As he admitted, there could be unintended meanings, and some of them can be existentialist or psychoanalytic ones.

[15] Cf. Smith 2010. Furthermore, Smith appropriately notes that existentialism and Kaila's philosophy are not mutually exclusive.

In this regard, it seems that Wartenberg's version, the "creator-oriented" approach which Livingston calls a conditional intentionalism, would more appropriately account for Bergman's broader influences. Moreover, despite Livingston defending a partial intentionalism, an emphasis is placed on the importance of investigating external sources to the work itself to the point where they are the sole reference. He acknowledges that there are cases where external evidence may be lacking, and that in such cases we can rely on internal evidence. Maybe he means that, in this case, the quality of the interpretation would be inferior. So, let us imagine that Bergman did not say anything about his works anywhere, or that only ambiguous or unclear statements were made about them: would it be impossible to interpret his films? Worse than that: would it mean that we could never affirm that his works philosophise?

Livingston means to argue that Bergman does not merely illustrate Kaila's philosophy, yet his arguments seem to imply just that. He says that Bergman is a filmmaker who was informed by and expressed philosophical thinking, but seems to rely on overlapping Kaila's philosophy and Bergman's work too much: the dichotomy established between Kaila's psychoanalytic influence shows just that. It is not clear why, in Livingston's approach, Bergman did not merely illustrate Kaila. If Livingston thinks that Bergman disagreed with Kaila on one thing (morality), why not think that he disagreed with more than that, and that he came across other ideas that could be contradictory to Kaila's ones?

Livingston's work on Bergman's philosophy shines a light on a part of Bergman's work that scholars were not familiar with. However, it seems to lack an understanding of how Bergman could have gone beyond Kaila, and how that could be part of his philosophical insight.

If, on one hand, we could agree that Kaila's philosophy book is a good explicit source to interpret Bergman's philosophy better, to limit our interpretations of his films as revolving around just a specific author or idea can be overly reductive, and we could miss other philosophical interests and insights that the author could have had.

Trying to make these partially intentionalist interpretations, which seem to rely so much on explicit intentions found in external sources to the films, is a strategy that Livingston adopted to, as previously observed, not only avoid the imposition objection, but also fulfil his idea of a bold thesis. It seems, however, that by attaching Bergman to Kaila in excess, whilst being indifferent to other aspects of his films, Livingston ends up by weakening Bergman's potential for defending a bold version of film as philosophy that combines both an appreciation of his films as works that belong

to the exclusive or paradigmatic cinematic medium, and an understanding of his philosophical original points of view.

Livingston charges philosophers' common readings of *The Matrix* for being extrapolative and not saying much about how the film is devoted primarily to "dazzling representations of fights, gun battles, and high-speed chases" (Livingston 2009, 198). We could also charge Livingston with focusing too much on Kaila's philosophical elements in Bergman's films. However, Livingston seems to try to say that philosophical interpretations of films do not count as adequate bases for film appreciation, since "one of the primary goals of the appreciation of a work of art is to understand and evaluate its artistry" (Ibid.). This tension in Livingston's approach highlights a broader challenge for understanding films as philosophical thought experiments. While we need some constraints to avoid pure imposition, overly restrictive intentionalism may prevent us from recognising how films can engage in genuine philosophical thinking through their distinctive cinematic means, even when that thinking extends beyond what the filmmaker explicitly intended.

In this account, Bergman can be seen as a filmmaker who philosophised through his works. The main advantage of Livingston's intentionalism is that, at least in terms of authorial intentions, we can say that Bergman intended to think philosophically in his films. We know that he was acquainted with some traditional philosophical texts and was imbued in a context that would allow him to have existentialist and psychoanalytic interests easily.

5.5 Implications for film as philosophy

Livingston's examination of Bergman's work demonstrates the risks of reducing a filmmaker's philosophical depth to a single influence. His analysis reveals the problems of an excessively constrained interpretation. By focusing on Kaila's influence, he risks dismissing alternative interpretations. As Dan Williams observes, there are clear overlaps between Kaila's ideas and those of other philosophers like Sartre (Williams 2015, 165), suggesting that philosophical influences are usually not singular or simple.

These limitations in Livingston's approach point toward the need for an alternative approach to the imposition objection that acknowledges the role of authorial intentions without falling into overly restrictive interpretations.

A problem that may emerge from viewing film as philosophical thought experiments, in particular, concerns what we might call the "reasoner challenge": who exactly is conducting the

philosophical reasoning, the filmmaker or the viewer? Traditional thought experiments typically have a clear reasoner who guides the reader through a line of reasoning. When Thomson presents the violinist hypothesis, she is the reasoner who constructs the thought experiments and guides readers through their implications. In film, the reasoning appears to be distributed between the filmmaker and the viewer. However, this apparent problem can be viewed as a strength, as cinematic thought experiments work through a collaboration between the filmmaker who establishes the parameters and conditions of the film, while viewers interpret it. We can recognise that cinematic thought experiments function as a collective and collaborative effort. The filmmaker is the "architect" of the thought experiments, establishing their conditions and guiding their development through cinematic techniques; the viewer serves as the "executor" of the thought experiment by employing those conditions to generate philosophical insights. The filmmaker constructs scenarios that test philosophical concepts through choices of framing, editing, sound design, and narrative; the viewer interprets and constructs mental models, evaluates competing interpretations and extracts philosophical implications from them. This requires the same kind of philosophical reasoning that traditional thought experiments require, but through different avenues.

Based on various intentionalist approaches, we can defend a moderate intentionalism that serves the analysis of cinematic thought experiments, with some essential principles:

1) Recognition of multiple philosophical influences: unlike Livingston's narrow focus on single sources, we can acknowledge that filmmakers exist within complex intellectual environments where multiple philosophical traditions may inform their work. A filmmaker like Bergman might be influenced by existentialism, psychoanalysis, and Kaila's psychology without reducing his work to just one source. The same can be said about professional philosophers who are influenced by several philosophical sources.

2) Emphasis on successful intentions: we focus on intentions that successfully "mesh" with what is presented in the work. This prevents both Humpty-Dumpty-ism, where anything goes, and overly restrictive interpretations.

3) Integration of internal and external evidence: We must examine both what films present directly and the broader contexts in which they were created. External evidence should illuminate interpretation.

4) Dynamic interpretation: following the "spectral author" concept, interpretation becomes a dialogue between the filmmaker's intention, textual evidence, and viewer engagement.

Moura demonstrates how films create what he terms a "spectral presence" that emerges from the work itself while retaining a connection to the actual director.

Building on Grice's theory of communicative meaning, we can understand cinematic thought experiments as instances of non-natural meaning that depend on recognisable intentions. However, unlike ordinary conversation, these intentions operate through cinematic techniques and cultural contexts. The filmmaker intends to create certain philosophical thinking through their work, expects viewers to recognise these intentions through their engagement with the film, and this recognition becomes part of the philosophical experience itself.

This resolution of the imposition objection and the reasoner challenge has implications for my broader thesis about film as philosophical thought experiments. By developing an understanding of how cinematic meaning emerges through the interactions of filmmaker intentions and viewer interpretation, we can address several points about film as philosophy through thought experimentation. First, when we understand filmmakers as philosophical agents who construct thought experiments through cinematic means, and viewers as philosophical agents who execute these experiments through interpretation, films can be viewed as authentic philosophy rather than a vessel for pre-existing ideas. Even when films present us with familiar philosophical concepts and problems, they do so in ways that generate new understandings through their distinctive modes of presentation. So, the collaboration between filmmaker and viewer, mediated by cinematic techniques, creates opportunities for philosophical thought that differ from what can be derived from textual and imaginative means. This perspective offers some guidance for analysing films as philosophical thought experiments: rather than seeking single and definitive interpretations based on discoverable authorial intentions, we should engage in a process of reflective equilibrium that considers multiple sources of evidence while remaining open to the philosophical possibilities that emerge through interpretation. This means iterating between, on one hand, particular interpretations supported by textual evidence and, on the other hand, general principles about what filmmakers could reasonably intend given their historical and cultural context. When conflicts emerge between textual evidence and our assumptions about authorial possibilities, we adjust both, refining our interpretive claims while also reconsidering what intentions are plausible. This reciprocal adjustment between particular cases (specific films) and general principles (authorial intentionality) continues until we reach coherence across the totality of available evidence.

Here, the imposition objection becomes an invitation to develop approaches to cinematic interpretation, rather than undermining film as philosophy. The resolution of this objection is not

to avoid philosophical interpretation, but to establish a direction for carefully investigating authorial intentions, recognising the fact that filmmakers may have multiple philosophical influences, while balancing respect for the filmmaker's creative vision with interpretation. The question is no longer whether films can philosophise by being philosophical and cinematic thought experiments, but how we can become better equipped to read and recognise the philosophical concepts that the films address.

We may also note that we can consider the philosophical credentials of many filmmakers when engaging in film interpretation. Directors such as Terrence Malick, Ingmar Bergman, Ethan Coen, Greta Gerwig, Michael Haneke, Wes Craven or Wes Anderson have formal philosophical education, yet chose cinema as their primary medium of expression. In cases such as these, this choice reflects not a fundamental division between philosophy and filmmaking, but a circumstantial one: a matter of professional focus rather than an intellectual chasm. Just as career philosophers produce traditional philosophical works such as books or articles, these filmmakers, by profession, can channel their philosophical ideas through film. The division can be viewed as a pragmatic choice in life, not a conceptual rift. Both domains require specialised training, time, and resources to achieve excellence. By circumstance, it may be impossible for a philosopher to produce a good example of a film that philosophises, or for a filmmaker to produce a good example of traditional philosophical work.

Furthermore, contemporary philosophical production involves multiple voices and interpretations as well. We do not study Kant in isolation, for instance. We also study his work through commentators, translators, and contemporary interpreters. This does not diminish Kant's contribution; we may argue the opposite, as it may be seen as being so rich that it invites interpretation, and it shows that philosophical thinking develops through ongoing dialogue. When philosophically trained filmmakers engage with philosophical concepts through film, they may be participating in this tradition of philosophical discourse through their chosen medium.

This chapter examined the imposition objection according to which philosophers merely project or impose their own ideas onto films rather than recognising genuine philosophical content within them. By grounding my analysis in Grice's theory of communicative meaning and exploring various intentionalist approaches to interpretation, I developed an understanding of how film can engage in philosophy without philosophers having to impose their ideas.

Wartenberg's creator-oriented approach is valuable for constraining interpretation within the bounds of what filmmakers could reasonably have intended, preventing arbitrary readings while respecting the cultural and intellectual contexts in which films are created. This approach is especially useful in avoiding anachronistic interpretations that attribute to filmmakers meanings they could not possibly have intended, given their historical and cultural circumstances, though the audience-oriented approach still preserves liberty of interpretation, even if anachronistic. Livingston's partial intentionalism reveals some limitations of overly restrictive interpretations, as demonstrated through his narrow focus on Kaila's influence in Bergman's work, which risks dismissing valid alternative interpretations.

The "reasoner challenge" – questioning who conducts philosophical reasoning in cinematic thought experiments – dissolves when we recognise that such experiments operate through collaborative philosophical discourse. The filmmaker serves as the architect, establishing the conditions through cinematic techniques, while viewers function as executors, engaging in interpretation. This collaboration enhances rather than undermines the philosophical value of cinematic thought experiments.

My moderate inflationism incorporates four principles: recognition of multiple philosophical influences, emphasis on successful intentions that "mesh" with the work, integration of internal and external evidence, and dynamic interpretation informed by the "spectral author" concept. This approach acknowledges that cinematic meaning emerges through the interaction between the filmmaker's intention and the viewer's interpretation, creating opportunities for philosophical engagement that differ from purely textual means.

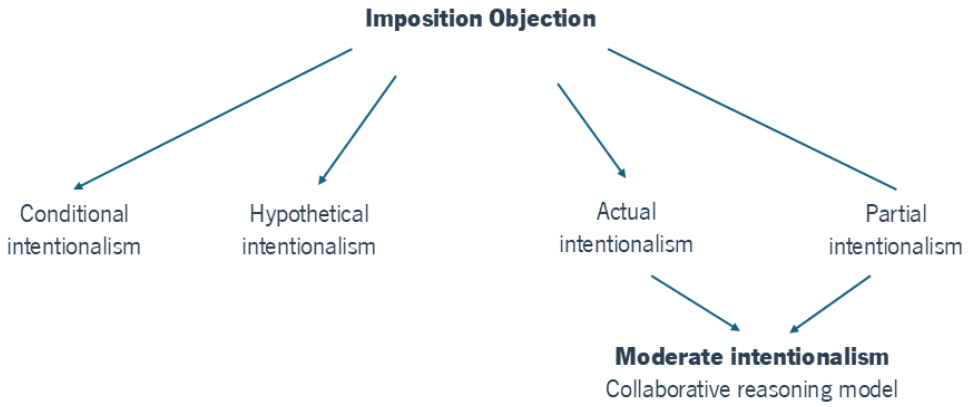
The imposition objection transforms into an invitation to develop approaches to cinematic interpretation. The objection gains further resolution when we consider that many contemporary filmmakers possess formal philosophical training, choosing cinema as their primary intellectual medium not from conceptual limitation but from professional focus. Just as philosophical thinking develops through ongoing dialogue involving multiple voices and interpretations, philosophically trained filmmakers participate in this tradition through their chosen medium.

Rather than asking whether films can philosophise, we should focus on something different: developing the interpretive tools we need to recognise and engage with philosophical work in films. This work emerges through a collaboration between filmmaker and viewer.

Having established this moderate intentionalist framework for cinematic interpretation, the following chapter will apply these foundations to demonstrate how specific films can engage in

philosophical thought experimentation. Drawing on the taxonomy of thought experiments established earlier, our understanding of their cognitive mechanisms, and the interpretive principles developed here, I will examine concrete examples of films that serve as counterexamples to philosophical theses, establish possibilities, demonstrate necessary connections, and contribute to philosophical discourse through their distinctive cinematic means.

Diagram 1: Intentionalist Responses to the Imposition Objection



6. Thought experiments on screen: case studies

6.1 Methodology and case selection

After arguing that films can function as philosophical thought experiments, I must address some challenges. I need to be clear about which films actually do this. Not every story counts as a philosophical thought experiment; otherwise, the concept loses its meaning. So let us establish what makes a film philosophical in this specific way.

Firstly, the film has to be able to be interpreted within philosophical conversations, whether by contributing to established debates or by developing concepts that can be connected to philosophical traditions. Here I should note something important: there can be an argument that films create entirely new philosophical concepts; however, we should question if those concepts really are new, and even if they are, they do not emerge in complete isolation from philosophical discourse. Even in philosophical practice, novel contributions build incrementally on established traditions, extending and recombining existing ideas in new ways. For us to recognise new conceptual contributions from film, we would need to demonstrate their connections to philosophical methods and concerns; otherwise, how could we identify such concepts as philosophical rather than any other form of epistemological contribution? This caveat does not prevent the possibility that films might contribute fresh perspectives or develop concepts in innovative ways. Instead, it acknowledges that philosophical progress *typically* occurs through incremental development, and that even groundbreaking contributions must engage with existing philosophical discourse to be recognisable as philosophical work. For example, let us consider structural films, which were philosophically innovative in that they reflected on the nature of film itself, often before such themes became common in mainstream philosophical discourse. Even in this case, the questioning of the nature of an art form is not something radically new; it represents a very traditional philosophical endeavour. This does not mean we should dismiss structural films' contributions to philosophy of film; quite the contrary. It just means that even this philosophical treatment of the nature of film common in structural cinema refers to broader philosophical concepts and discussions, and this is how we identify the philosophy present in them, Thus, even if one could argue that film, or even philosophy, creates radically new philosophical concepts, that is not the argument we are making, nor one we need to make. So, for instance, when Wartenberg argues that *The Matrix* poses the deception hypothesis, he relates that to Descartes' evil genius.

This strategy of relating a concept found in film to traditionally known philosophical concepts is not only pragmatic, but it may also show that such topics are repeated throughout history and forms of expression.

The film also must not just employ narrative structure in general (it may not even have a narrative structure at all, such as the case for structural films), but it should mirror the methodological structure of philosophical thought experimentation. This means presenting counterfactual scenarios designed to test specific philosophical intuitions, theories, or concepts. The film should function as a controlled thought experiment that isolates particular variables for philosophical examination. For example, *Eternal Sunshine of the Spotless Mind* does not just explore memory and relationships in general: it presents the scenario "What if we could selectively erase painful memories?" and then explores the philosophical implications through Joel's experience. The film's structure mirrors the logical progression of a philosophical thought experiment. Similarly, a non-narrative film can have the same kind of structure. For example, *Serene Velocity* presents the scenario "What if we stripped cinema down to its most basic elements?" It then isolates the variables like focal length, framing and temporal rhythm to test our intuitions about what constitutes cinema. By shifting between two different zoom positions in a corridor, it examines whether narrative, character or representational content is necessary for cinema, and if not, what are the necessary elements.

Finally, the film must be structured so that philosophical understanding can emerge through the viewing experience itself. The film should invite philosophical interpretation without requiring specialised training to access its philosophical content, although such training helps access the possible deeper philosophical connections and articulate them.

In summary, we should see if the film we are examining:

a) Is interpretable within philosophical conversation;
b) Follows the structure of a philosophical thought experiment;
c) The philosophical understanding emerges through the cinematic experience itself.

In this chapter, I will examine how specific films function as different types of philosophical thought experiments. My analysis is based on the moderate intentionalist approaches, which respect authorial intention while recognising how cinematic meaning emerges through the interaction between filmmaker and viewer. Thus, following my resolution of the imposition objection, we recognise that films can philosophise through the collaboration of filmmakers who

establish the conditions of the thought experiment through cinematic techniques, and viewers who interpret them.

This approach is guided by several important methodological principles established earlier:

1) Moderate inflationist account of thought experiments: I apply the moderate inflationist understanding of thought experiments, especially the mental modelling and constructivist accounts, in order to examine how films can activate cognitive resources and reconfigure conceptual understanding through quasi-observation;
2) Focus on cinema's characteristic techniques: I apply a flexible understanding of medium specificity that avoids essentialism and focus on cinema's characteristic techniques such as editing, framing, sound design, temporal manipulation and so on;
3) Applying the taxonomy of thought experiments: I examine how films can function as specific types of thought experiments and which types are particularly well-suited to cinematic presentation;
4) Moderate intentionalism: I recognise that in film interpretation, there is collaborative reasoning. The filmmaker creates a thought experiment in cinematic form, establishing its conditions through cinematic techniques, while viewers interpret it.

This approach directly addresses the major objections to film as philosophy examined in previous chapters. By focusing on how films activate cognitive resources through mental modelling, we move beyond concerns about explicitness: films do not need to present formal arguments to philosophise. The collaborative nature of cinematic thought experimentation addresses the generality objection, as viewers extrapolate from particular cinematic scenarios to broader philosophical implications. Finally, my moderate intentionalist approach avoids the imposition objection by grounding interpretation in the interaction between the filmmaker's intentions and cinematic evidence.

We here focus primarily on the work of Wes Craven, known for his popular horror films. Craven had a formal academic training in both psychology and philosophy, earning an undergraduate degree in English and Psychology from Wheaton College in 1963, followed by a master's degree in Philosophy and Writing from Johns Hopkins University in 1964. This matters because his films show this training. His psychology background shows up in how he explores consciousness and mental states, while his philosophy education appears in how he tackles questions about knowledge, reality, and human nature.

However, despite this, Craven has been overlooked, primarily by philosophers interested in the philosophy of film in general, but more importantly by those interested in film as philosophy in particular. This reflects unexamined assumptions about the relationship between popular culture and serious philosophical discourse. However, I should note that some philosophers have already demonstrated horror's philosophical interest. Carroll has established this through cognitive and aesthetic analysis in *The Philosophy of Horror* (1990), while Freeland has explored horror's engagement with fundamental questions about humanity, gender, and morality in *The Naked and the Undead* (2000). My approach takes a different direction. I examine how horror films function as philosophical thought experiments rather than serving as objects for philosophical analysis.

There seems to exist a tendency to dismiss popular Hollywood cinema as a whole, and specifically horror films as mere "chills and thrills", which obscures the philosophical work accomplished within these kinds of popular genres. The contrast between Craven's philosophical training and his reputation as a horror filmmaker who created some of the most shocking and controversial films may illustrate how effective philosophical work can occur even in such a context. If good philosophical thought experimentation can be identified in films commonly regarded as lowbrow entertainment, then the move from these most challenging cases to more readily accepted ones should be less controversial.

It is well known among Craven's fans and scholars that his strict fundamentalist Baptist upbringing forbade him from watching films, leading to his first cinematic experiences during college. This later exposure to cinema, particularly Ingmar Bergman's *The Virgin Spring* (1960), would inspire him to create his first, and one of the most controversial, films: *The Last House on the Left* (1972). As he stated in an interview, "I very consciously used the same medieval morality tale about a father taking revenge on the shepherds who raped his daughter" (Craven 1984). Craven's projects were explicitly connected to contemporary American violence and the Vietnam War. As he explained,

> I made *Last House* when I was examining the whole sickness associated with Vietnam. It was fascinating to take characters who had killed, who'd done something utterly reprehensible, and make them worry about being dressed properly for dinner at the house or who was going to sleep with whom. It was a strange turnaround that really snapped people's heads. Western culture has traditionally dealt with barbarism by projecting it outward. There's that old rationalization that we're only dealing with animals, so we have the right – or God has given us the right – to be animals and kill the Aztecs, kill the Indians,

> enslave the Blacks, drop bombs on the Japanese. One of the healthier notes of the latter twentieth century is that Americans are finally looking in the mirror and seeing that the enemy is in us as well as in others, that there really is no essential difference between the enemy and us. We all have the capacity to be maniacal or go crazy (Craven 1990).

What we may now call Craven's philosophical project spans four decades and addresses mainly philosophical questions about the nature of reality, violence, moral responsibility and the role of religion in action. His "recurring themes include religion, the nature of reality, the destruction of the American family and more" (Catt 2022), while he was "always interested in analyzing why things scared people and really exposed the cultural fears and anxieties in horror" (Johnson and Hayes 2023). His career demonstrates that while he maintained consistent philosophical interests, they evolved, deepened, and refined over time. The philosophical nature of Craven's work also becomes evident when considering his explicit statements about the concepts in his films, which will be further explored when analysing specific films. I will also analyse whether his statements coincide with what the films show.

This chapter is organised around four major case studies that demonstrate different types of philosophical thought experiments in Craven's work. First, I consider how *The Last House on the Left* serves as a counterexample to conventional moral intuitions about revenge. Second, I analyse how *A Nightmare on Elm Street* (1984) establishes philosophical possibilities regarding the instability of boundaries between dreaming and waking consciousness and reality. Third, I examine how *Scream* (1996) functions as metacinematic commentary, counterexemplifying and establishing possibilities about the effects of media on violence and of hyperreality and simulation. Finally, I analyse *Wes Craven's New Nightmare* (1994) as Craven's culminating philosophical statement, a meta-textual work that synthesises his longstanding concerns with violence, reality and representation while demonstrating the consistency and evolution of his philosophical project across decades.

These four films are clear expressions of Craven's philosophical thought experimentation; I chose them for their explicit engagement with fundamental philosophical questions rather than their commercial success or cultural impact, which are also undeniable. While other significant works in Craven's filmography – including *The Hills Have Eyes* (1977), *Shocker* (1989) and *The People Under the Stairs* (1991) – contain philosophical elements similar to the films I am analysing, the selected ones most clearly demonstrate the methodological principles I have established: they employ distinctively cinematic techniques to thought experimentation that would be difficult to

replicate through purely textual means. Moreover, these works span Craven's career since his 1972 debut, allowing us to trace the development and refinement of his philosophical concerns over time.

While my primary focus is on Craven's work, it is worth noting that the philosophical thought experiments I identify in his films are not unique to his oeuvre. Other filmmakers, working independently and often without philosophical training, have created works that engaged with similar subjects and in similar ways.

Through this analysis, I also aim to demonstrate that the relationship between film and philosophical thought experimentation is not uniform. Some types of philosophical thought experiments may prove more amenable to the cinematic medium than others, and even individual films vary in effectiveness and epistemic power.

6.2 *The Last House on the Left*: a counterexample to common moral intuitions

The Last House on the Left functions as a philosophical thought experiment, specifically a counterexample to conventional moral intuitions about revenge. Released in 1972, Wes Craven's debut feature follows a narrative structure that initially appears to be simple. It centres on two teenage girls, Mari Collingwood and Phyllis Stone, who go into the city to attend a concert. They encounter a group of escaped convicts led by Krug Stillo (David Hess), consisting of his son Junior, the sadistic Weasel, and Sadie, a woman complicit in their crimes. The criminals kidnap the girls, subjecting them to prolonged sexual torture before ultimately murdering them in the woods. Later, the criminals seek shelter at the Collingwood family home, posing as travelling salespeople. Mari's parents, Dr. John and Estelle Collingwood, discover the truth about their guests' identities and the fate of their daughter. The final act depicts the parents' methodical revenge against their daughter's killers, employing brutal methods of execution.

The Last House on the Left belongs to the rape-revenge subgenre of exploitation cinema, which commonly follows a three-act structure: sexual assault, aftermath and violent revenge. This genre, exemplified by films such as *The Virgin Spring* (which, as we have seen, directly inspired Craven's work), *I Spit on Your Grave* (1978), *Ms. 45* (1981), or *Promising Young Woman* (2020), usually invites sympathy for vigilante justice by presenting sexual violence as the ultimate moral transgression that justifies extreme retribution. Such films characteristically eliminate legal

alternatives (for example, by presenting incompetent or dismissive law enforcement), making them good vehicles for testing audience intuitions about the ethics of revenge.

6.2.1 Targeting revenge theory

The film can be seen as primarily functioning as what Sorensen categorised as a "necessity refuter", or a counterexample. Specifically, the film challenges commonly held moral intuitions about the nature of civilised behaviour and the proportional response to injustice, especially the tendency to respond in a vengeful manner. Its most direct philosophical target is a philosophical tradition that defends revenge as a morally adequate response to serious wrongdoing.[16]

The foundation for the moral legitimacy of revenge is found in multiple philosophical sources that offer justifications for why revenge preserves, rather than corrupts, moral balance. Aristotle's analysis in the *Rhetoric* provides some understanding of justified anger and revenge. He defines anger as a "a longing, accompanied by pain, for a real or apparent revenge for a real or apparent slight, affecting a man himself or one of his friends, when such a slight is undeserved" (Rhetoric II.2.) According to Krisanna Scheiter's interpretation of Aristotle, he viewed revenge as serving "a purpose in restoring balance and justice" where "the main purpose of revenge for Aristotle is to restore justice and ensure that one is not mistreated by the offender in the future" (Scheiter 2010, 1).

Aristotle's definition establishes that justified anger must target specific individuals rather than expressing general hostility: "the angry man must always be angry with a particular individual (for instance, with Cleon, but not with men generally), and because this individual has done, or was on the point of doing, something against him or one of his friends" (Rhetoric II.2.2). According to Scheiter, Aristotelian revenge requires three elements: the wrongdoer must have intentionally and undeservedly slighted the victim; the wrongdoer must be aware that the victim is causing their suffering; and the wrongdoer must understand they suffer because of their original transgression (Scheiter, 3-4). Aristotle emphasises this recognition requirement, noting that "anger has to do with the individual" and requires that wrongdoers know "that the punishment comes from them in

[16] This analysis focused on personal revenge rather than institutional retributive justice, though both involve responses to wrongdoing. Nozick argues that revenge and retribution are distinct concepts: revenge is inherently personal and involves "pleasure at justice in the suffering of another", while retribution can be impersonal and involve "pleasure at justice being done" (Nozick 1981, 367). George Fletcher similarly emphasises that retributivism "is not to be identified with vengeance or revenge, any more than love is to be identified with lust" (Fletcher 2000, 417). *The Last House on the Left* specifically examines the moral psychology of personal vengeance rather than questions of institutional punishment, though the film's implications may extend to both domains.

requital for their own wrongs" (Rhetoric II.3.16). This structure preserves moral distinction by requiring that revenge respond to prior wrongdoing rather than expressing anger arbitrarily.

The common intuitions regarding the morality of revenge and the argument defending revenge's moral adequacy follow this kind of reasoning:

1) Serious wrongdoing creates a moral imbalance that demands correction;

2) Revenge serves as a legitimate moral correction by reasserting the victim's dignity and moral boundaries;

3) Those pursuing justified revenge maintain moral legitimacy because they respond to rather than initiate wrongdoing;

4) The moral distinction between aggressor and defender remains intact when defence takes the form of proportional revenge;

5) Therefore, revenge is a morally adequate response that preserves the victim's moral standing.

This position can also be supported by just war theory and self-defence doctrine, maintaining that violent responses to aggression are morally distinct from aggressive violence itself. However, the case found in *The Last House on the Left* is limited to personal revenge perpetrated by the victim's parents.

6.2.2 The film's counterexample structure

Now, as said above, *The Last House on the Left* presents a counterexample to this entire thinking, by following this reasoning[17]:

1) The Collingwoods suffer the most extreme wrongdoing imaginable – the sexual torture and murder of their innocent daughter;
2) They pursue revenge against identifiable wrongdoers who have escaped legal justice;
3) Their revenge employs methods as brutal and sadistic as those used by the original criminals;
4) The film presents the parents as morally equivalent to the criminals through identical visual framing and emotional emptiness;
5) Therefore, even maximally justified revenge corrupts victims into moral equivalents of their victimisers and does not bring any sense of justice.

[17] By reconstructing the film's reasoning, we are not stating that it is an argument, as thought experiments are not arguments either.

The film undermines each element of the pro-revenge position. Against Aristotelian requirements that revenge must target specific wrongdoers who understand why they suffer, the film demonstrates that even when these condition is met – the Collingwoods target the exact individuals who killed their daughter, and the criminals clearly understand they suffer from their crimes – the apparent psychological satisfaction of revenge transforms the avengers into sadistic killers.

However, we might note that this initial formulation risks reducing Aristotelian revenge theory to a weaker position than it defends. Aristotle would not advocate for sadistic revenge, and his framework explicitly sets moral bounds on justified anger and retribution. We can refine the argument to engage with revenge theory on its stronger terms. While Aristotelian revenge theory sets moral bounds on justified revenge, *The Last House on the Left* challenges whether such bounds can be maintained in practice. The film shows us the Collingwoods beginning with justified anger directed at specific wrongdoers. Yet as we follow their journey, we witness how the psychology of revenge escalates beyond any moral constraints. We see these justified avengers transform into something else entirely: sadistic killers who mirror the very people they sought to punish.

This shifts the philosophical focus from whether revenge can be theoretically justified to whether it can be practically contained within moral limits once it has been initiated. This interpretation shows us that the film challenges a particular aspect of revenge. As we have seen, it does not question the theoretical possibility of revenge as a moral response. Instead, it challenges whether we can practically sustain moral constraints once we enter the psychology of vengeance.

The counterexample becomes more interesting as it engages with the revenge theory on its terms, while revealing inherent limitations in the containment of vengeful impulses – something that could be seen as an advantage over traditional textual thought experiments, as the film is emotionally charged, detailed, and closer to our everyday lives.

So, now I can specify how exactly this film counterexemplifies pro-revenge positions:

1) Revenge can be a morally adequate response to serious wrongdoing when it operates within proper moral bounds – targeting specific wrongdoers, maintaining proportional response, and ensuring recognition requirements – preserving the moral distinction between justified retribution and arbitrary violence;

2) *The Last House on the Left* presents the Collingwoods as beginning with properly directed anger that meets all Aristotelian requirements, yet their pursuit of revenge psychologically escalates beyond any theoretical constraints, transforming them into

sadistic killers indistinguishable from their daughter's murderers despite starting within established moral parameters;

3) The film, thus, shows that the psychology of vengeance is unpredictable and can escalate beyond any theoretical moral bounds;

4) Therefore, revenge cannot serve as a morally adequate response, even when it operates within proper moral bounds.

6.2.3 Cinematic techniques

The film's philosophical impact depends on specific cinematic techniques that enhance its function as a thought experiment. The film's most significant technique is its documentary-like visual style, achieved through the use of a handheld 16mm camera that creates "the allure of the found footage aesthetic as we know it today" (Flasz 2020).



Figure 21: Poster advertising *The Last House on the Left*.

This aesthetic choice does something important, as it refuses to distance us in the way that traditional cinematic styles often do: "by sweeping, panning and tilting the camera in the most chaotic fashion, and even putting it on the ground for it to capture this harrowing ordeal from a rather unusual perspective, he [Craven] created an illusion that what the audiences were watching was a documentary or a snuff film" (Ibid). Craven was explicit about his intention to create visceral realism:

> I genuinely felt I portrayed what it actually felt like to kill somebody for real. The killing is absolutely heartless, the protracted violence was very human to me. I wanted to make a statement about violence and American movies at the time. You must remember that I made the film at the time when people were watching villages being burned on the news as they ate their dinner. Nobody comes out of *The Last House on the Left* looking noble after "they have killed somebody, even though it may have been for the right reasons" (Craven 1982).

This technique activates the "quasi-observational" processes present in thought experiments. The handheld aesthetic creates an embodied viewing experience: the camera's instability mimics how human vision works during moments of stress, creating physical tension that mirrors the character's experience. The unstable framing prevents viewers from comfortable observation; instead of just watching violence, we feel the proximity to it. While reading, we maintain a controlled pace (we can pause or skip ahead), engage in abstract mediation (the text requires mental translation into images), and generate self-regulated imagery (we control the intensity of our mental pictures). The handheld cinematography employs a different process: we see blood, not the word "blood"; the film's temporal structure dictates the pacing, and we involuntarily pay attention to specific visual details that the camera reveals. When viewers watch the Collingwoods' revenge sequence through shaky, intimate camerawork, they cannot maintain a distance that allows for a "clean" moral thinking. This is why viewers report feeling "dirty" or morally complicit after watching the film, as if we were watching real footage.

This emotional engagement shows how film thought experiments work better than traditional ones. Critics like Machery say thought experiments become unreliable when they go beyond the proper domains of our psychological capabilities, as I previously explored. But the film shows moral decisions in the rich emotional contexts where they actually happen. A specific subplot

also serves an important function in this thought experiment: the bumbling police subplot. It involves two police officers who are portrayed in a slapstick, comedic manner throughout the film.

They often get lost while trying to respond to emergencies, experience car trouble, and fail to provide any effective law enforcement presence.



Figure 22: Still image of deputy Harry and Sheriff Marshall Anker, the two bumbling police officers.

This incompetence reflects what Carroll identifies as a typical horror trope: the failure of authority figures, including law enforcement, to provide protection when supernatural or extreme threats emerge. Carroll specifically notes that in horror narratives "the discovery that a monster is at the root of recent evil is resisted, often by the powers that be," with "authority figures such as the police, eminent scientists, religious leaders, government officials, or the army" treating credible threats with scepticism (Carroll 1990, 101). While *The Last House on the Left* deals with human rather than supernatural monsters, it follows this same pattern of authority failure. The sheriff's incompetence mirrors Carroll's observation about how horror plots show the failure of authority to protect, forcing protagonists to confront threats without official support. This type of narrative serves several purposes: it removes the safety net that authorities are supposed to provide while simultaneously creating what Carroll calls a "delightfully superior position" for the audience, who can see the threat clearly (Ibid., 102). In this film, Craven's audience experiences mounting frustration as the sheriff proves incapable of preventing the unfolding violence, ultimately justifying the parents' vigilante response as the only viable solution when official channels fail.

This subplot is tonally inconsistent with the rest of the film, as the comedy contrasts with the brutal and horrific main narrative, which also serves an important function, as the film eliminates institutional justice as a viable alternative to personal revenge. The complete failure of legal authorities initially appears to strengthen the case for the Collingwoods' revenge: if the proper channels of justice are absent, does this not somehow justify taking justice into their own hands? For example, Francis Bacon acknowledges that while revenge is generally "wild justice" that law should "weed out", it becomes more tolerable when "there is no law to remedy" the wrong (Bacon 1625).

However, this setup makes the film's counterexample more powerful. By removing any alternative to revenge and creating conditions where personal vengeance appears most justified, the film tests revenge under the most favourable conditions. When institutions fail, personal revenge becomes the sole option; yet, the film demonstrates how this leads to a moral equivalence between the victim and the victimiser.

A particularly significant sequence occurs after the murders when Krug's gang stands silently watching Mary's body disappear into the lake.





Figure 23: Sequence of Krug's gang standing silently after Mary's murder.

Here, “the expression on their faces seems to acknowledge that they know the last threads of their humanity are following her to the grave” (Chaka 2020). This moment complicates the categorisation of criminals as purely evil, as another common thread I find in Craven’s films is the moral ambiguity that anticipates more contemporary horror. This questioning of simplistic moral categories would become increasingly prominent in horror cinema.

The climactic revenge sequence employs specific cinematic techniques to emphasise the moral transformation of the parents, too. The camera’s positioning during Dr. Collingwood’s methodical torture of Krug creates a perspective that refuses to provide a comfortable distance from the action. Similarly, the film does not employ any heroic or triumphant music or cinematography during this sequence, which denies the usual cathartic satisfaction associated with justice being served. Instead, the almost clinical presentation emphasises the parents’ moral degradation.



Figure 24: Still image of Dr. Collingwood’s methodical torture.

6.2.4 Collaborative reasoning

Understanding this film’s philosophical content, as a thought experiment, requires examining the relationship between the film’s internal evidence and Craven’s explicit intentions. Craven statements about “how the Christian parents in the film are brought to an animalistic rage, which is no doubt an expression of Craven’s rejection of his own religious background” (Housman 2020) provide external evidence of philosophical intent. However, this intention succeeds philosophically

only because it meshes with the film's internal evidence. The systematic visual parallel identified by Muir (2004, 50-51), the elimination of institutional alternatives through the police subplot, and the presentation of violence all support his stated intentions to challenge conventional moral categories. This also suggests the film's awareness of its challenge to conventional moral categories and its exploration of circumstances that might corrupt apparently civilised individuals. Craven's intention to expose the moral equivalence between civilised and savage was deliberate:

> In *Last House* I was trying to relate to that side of us that we like to externalize, the guy in the black hat. I tried to find out whether that was coming from my own soul, how it could get so loose, how it could happen like that. And the thing that seemed to be was (without getting into pat psychological explanations) that those terribly violent parts of humanity which we like to externalize – especially with Americans who say, "They were the Indians, Russians, dope addicts, blacks" – were us. They were in all of us, part of our fibre. Although we were a very civilized nation, it was really a hollow lie (Craven 1982).

The collaborative nature of cinematic thought experiments is evident here: Craven architects the philosophical conditions through cinematic techniques, while viewers execute the thought experiment and its extension (i.e. its examination and broader implications). This collaboration requires viewers to perform specific cognitive work that differs from passive film consumption. First, viewers must actively resist their natural sympathetic alignment with the grieving parents and recognise the moral equivalences the film establishes through visual paralleling and narrative structure. Traditional textual thought experiments require us to construct mental simulations ourselves. However, when we watch films, we find that they externalise aspects of this process through their presentation, in this case, by recognising that the parents' transformation into killers demonstrates what happens when revenge is pursued under seemingly justified conditions. Second, viewers must maintain enough emotional engagement to access the embodied aspects of moral reasoning while simultaneously developing sufficient distance to process the thought experiments' implications. This requires experiencing the visceral impact of violence while simultaneously analysing what revenge actually entails. Viewers must recognise this scenario as testing universal claims about the moral adequacy of revenge.

While viewers with philosophical training may more readily identify the specific ideas being tested, recognising Aristotelian revenge theory or understanding how counterexamples function in philosophical methodology, this cognitive work remains accessible to general audiences.

Philosophical training provides advantages in explicitly articulating what moral principles are at stake and connecting the film's scenario to broader debates, but the essential thought experimental work does not require specialised knowledge. The film presents strong emotions that allow viewers to connect with the moral issues in the story, even if they are not familiar with philosophical terms. This is one of the unique ways that film contributes to philosophical conversations. A viewer does not need to understand Aristotle's ideas about justified revenge to see that the Collingwoods start with valid complaints but end up as morally questionable killers.

The film's relationship to *The Virgin Spring* provides additional context. Craven explicitly acknowledged the source: "Yeah, Bergman's *The Virgin Spring* was, you know, obviously the basis of *The Last House on the Left*. I always say that *Virgin Spring* was based on a medieval... they were called lays – not what they sound like – but they were songs with stories, so it's a pretty ancient kind of metaphor for getting the experience of life the hard way" (Craven 1999). While both films explore violence and revenge, Craven explicitly removes the religious framework, creating a "godless world" where characters exist as "realistic people caught in a universe of purposeless, random violence" (Muir 2002, 211).

This kind of secularisation transforms what could be a meditation on divine justice into a purely human moral experiment, eliminating transcendent resolution and forcing audiences to confront the problem without supernatural or institutional consolation.

Some critics have questioned the film's philosophical coherence, particularly its apparent equation of victims with victimisers: "When Craven equates victims with victimisers in these films, I don't see the morality. Or rather, I see only the absolute morality of abstract philosophy, a morality divorced from real life as it is lived and experienced by real human beings in the real world" (Callahan 2020). Callahan further argues that this approach "comes off as pretentious, self-righteous, and sanctimonious" (Ibid). This critique raises an important question about whether the film's moral conclusions are too removed from lived human experience to be meaningful. Interestingly, in this way, this critique mirrors the criticism made against the use of thought experiments in philosophy, as it also seems to be an accusation of the film being too philosophical. Now, this accusation may point to one of its philosophical strengths: it is effective as a moral thought experiment, which activates our "intuition pumps", by making us uncomfortable by challenging our common assumptions. The potential that the film entails controversial philosophical ideas, at least according to some critics, does not invalidate it as a good philosophical thought experiment. The film's effectiveness as a counterexample does not depend on whether

audiences accept its apparent moral equivalences, but on whether it successfully challenges the pro-revenge thesis. The continued debate about the film's moral implications forty years after its release suggests that it remains an active philosophical provocation.

The Last House on the Left is not unique in its challenge to pro-revenge positions. Other horror films can be seen as critical of revenge as morally adequate. For example, *Carrie* (1976) presents revenge as an indiscriminate force that cannot distinguish between guilty and innocent targets. Another excellent example of such a film is *Pumpkinhead* (1988), directed by Stan Wilson.

Pumpkinhead shows the story of Ed Harley and his seek for revenge. After his young son is killed in a hit-and-run accident by a group of careless city teens who try to escape the area without any witnesses, Ed, a grieving father living in a rural area, seeks revenge. He turns to a local witch who helps him summon a legendary demon known as Pumpkinhead, a monster that exacts vengeance on those who have wronged the summoner. As the monster begins hunting down the responsible teens one by one, Ed starts to feel a psychic connection to the creature and begins to realise the cost of his vengeance. Full of guilt, he tries to stop the monster, but the only way to stop it is by killing himself.

The film poses the question: "What if revenge were embodied as an unstoppable supernatural force?" Ed Harley's initial motivation appears to align with the same pro-revenge position addressed earlier. In the case of *Pumpkinhead*, the wrongdoers (the teenagers) have slighted Ed through their reckless behaviour that killed his son; they are aware of the harm they have caused, and Ed's desire for revenge stems from a genuine sense of injustice rather than arbitrary malice. However, the film demonstrates how even this seemingly justified revenge can spiral out of control. The monster starts killing everyone involved in the accident that ended up killing the son, then eventually he even kills characters who did not directly cause it and tried to help, and just characters trying to protect the perpetrators.

The creature itself is a literal, concrete and visible incarnation of the concept of revenge. It embodies the potential unstoppable nature of revenge: there is no possibility of stopping it, as nothing but anger drives it. There is no mercy or forgiveness. This theme of uncontrollable revenge is far from new. Aeschylus's *Oresteia* presents perhaps the first major artistic formulation of this problem, depicting how each act of vengeance begets another, trapping characters in an endless cycle of bloodshed. Pumpkinhead updates this ancient understanding for contemporary horror, literalising the supernatural force that perpetuates the cycle. There is a church sequence where two teens try to hide in an abandoned Christian church, invoking a tradition in which sacred spaces

are considered places of mercy and forgiveness. The monster's violation of this space, stepping into the church with no problem or hesitation, and even crushing a big cross as if it were made out of paper, demonstrates the incompatibility of revenge with core Christian values. The film's narrative can be reconstructed by developing the following argument:

1) Ed Harley suffers the death of his beloved son, which the legal system is unable to address;
2) In his grief, he seeks revenge by supernatural means, summoning Pumpkinhead to punish those responsible;
3) The monster kills the culprits brutally, revealing that vengeance cannot be precisely or ethically controlled;
4) As the killings progress, Ed becomes physically and morally entangled with the monster, experiencing its violence as his own;
5) Thus, even emotionally justified revenge unleashes forces beyond the avenger's control, ultimately transforming the victim into a conduit of destruction with no hope for justice or closure.

While *Pumpkinhead* and potentially other films engage with the same philosophical themes and may even be seen as counterexemplifying pro-revenge positions in general, *The Last House on the Left* is a particularly interesting candidate for a philosophical thought experiment for several reasons that go beyond Craven's philosophical education. Unlike films that explore philosophical themes, *The Last House on the Left* eliminates variables that might complicate the moral equation. The bumbling police are not just comedic relief; they serve the function of undermining institutional justice as a viable alternative. The film does not create its moral dilemma arbitrarily. The Collingwood's situation meets all criteria for justified revenge. This precision is important for the thought experimentation. The film's documentary style refuses the distancing effects of traditional cinematic style, as we have seen. The parallel framing of parents and criminals, the presentation of the revenge sequence, and the elimination of cathartic music or imagery serve the same purpose: to show the equivalence of victim and victimiser, and how revenge does not bring moral balance.

This parallel is done in several ways. From the opening scene, Craven creates rhymes between the Collingwoods and the Krugs. As Muir observes, the film opens with "John Collingwood sit[ting] in the living room of his home and read[ing] the newspaper, blissfully ignoring his wife," while "the camera later finds Krug's in their home, where the mother figure, Sadie, is also ignored by the men of the household" (Muir 2004, 50). These early visual parallels hint that the two families are mirror

images instead of moral opposites. The film's climactic visual strategy reinforces this equivalence through what Muir calls the "final connection between the two families" (Ibid). The concluding freeze-frame on the Collingwoods deliberately replicates "the earlier shot of Sadie, 'Weasel' and Krug at the lake after Mari's rape" where "the killers realise what they have done and they stand in the composition looking uncomfortable and ashamed, much as John and Estelle feel after dispatching their opponents in the finale" (Ibid., 51).









Figure 25: Still images: both families occupy the same visual and moral territory as killers who recognise their degradation.

This repetition visually shows that both groups occupy the same moral territory: both are killers who recognise their moral degradation. By framing both families identically in moments of violence and shame, the film does not preserve the moral distinction between the two.

The film does not attempt to prove that revenge is always wrong; instead, it presents one instance where even what seems to be justified revenge corrupts the avenger, thereby denying universal or general pro-revenge theses. The film appears to be explicitly designed to challenge philosophical assumptions, and its emotional impact is integral to the thought experimentation. By creating moral revulsion at both the initial crimes and the parents' revenge, the film successfully questions the moral standing of the avenger: viewers do not just think about the moral equivalence between victims and victimisers, they feel it, creating a form of moral understanding that purely abstract reasoning cannot achieve.

The film's effectiveness as a philosophical thought experiment also depends on how it activates mental modelling processes specifically. Traditional philosophical discussions of revenge operate through what we could call propositional mental models, i.e. abstract representations where we imagine moral principles applying to hypothetical scenarios. The film forces what we could call embodied mental modelling. When viewers experience the parents' transformation, they construct mental models that integrate not only logical relationships but also emotional and physical dimensions. The film activates "chucked material" from our experiences of violence, family relationships and moral degradation, but reorganises these elements in ways that reveal previously hidden connections. In this case, the film has a better empirical force than textual thought

experiments. Because it approximates the actual phenomenology of moral decision-making under extreme stress, the mental modelling here becomes embodied.

The analysis of this film may also reveal why counterexamples may represent the most effective type of thought experiment for cinematic presentation. As Bruce Russell's critique of *Crimes and Misdemeanors* demonstrates, films face more challenges when attempting to establish general philosophical theses – although, as I have argued before, thought experiments do not always have the purpose of establishing general theses. *The Last House on the Left* does not need to prove that revenge is always destructive, although it could be argued that it does that too. It only needs to demonstrate that revenge can be destructive under seemingly justified conditions, which it does through its visceral presentation. In this case, viewers experience the consequences of revenge through the psychological identification or sympathy with characters.

The counterexample type of thought experiment is the most successful at responding to the generality and explicitness objections: Gettier cases, for instance, do not provide explicit arguments either, presenting scenarios that challenge intuitions and forcing us to reconsider philosophical theses. *The Last House on the Left* presents what we might call a lived counterexample, one that engages our rational and emotional faculties simultaneously. Because the film offers no explicit commentary, it invites viewers to do the philosophical work themselves – not unprompted, but as a result of how Craven presents the story: as a collaborative exchange between filmmaker and audience.

6.3 *A Nightmare on Elm Street*: establishing the possibility of a perceptual drift

A Nightmare on Elm Street (1984) follows Nancy Thompson (Heather Langenkamp), a teenager in the suburban town of Springwood, who discovers that she and her friends are being stalked by a disfigured serial killer named Fred Krueger (Robert Englund), who attacks them through their dreams. When Krueger kills victims in their dreams, they die in reality as well. Nancy learns that Krueger was a child murderer who escaped legal punishment, prompting the parents of Elm Street to burn him alive in his boiler room hideout. Before his death, Krueger gained the supernatural ability to continue his killing spree through the dream world, targeting the children of those who killed him.

As Nancy's friends fall victim to Krueger one by one – first Tina Gray (Amanda Wyss) is killed in her sleep, then Rod Lane (Jsu Garcia) dies in his jail cell, and finally, Glen Lantz (Johnny Depp) is consumed by his own bed – Nancy realises that the adult authorities, including her police lieutenant father Donald Thompson (John Saxon) and her alcoholic mother Marge (Ronee Blakley), refuse to acknowledge the supernatural threat. Nancy must confront Krueger alone, eventually pulling him into the physical world where she can fight him on equal terms. The film concludes ambiguously, with Nancy appearing to defeat Krueger by withdrawing her fear from him, only for the final scene to suggest that the entire experience may have been an elaborate multilayered dream.

A Nightmare on Elm Street can be seen as functioning as a thought experiment that establishes a possibility, demonstrating the conceivability of several radical philosophical scenarios. As we have seen, Wartenberg notes some thought experiments "aim to conceptualise and establish the possibility of a scenario" (Wartenberg 2007, 60-61), arguing that *The Matrix* establishes the deception hypothesis, analogous to Descartes' evil demon. We may argue that *A Nightmare on Elm Street* does something similar.

The film primarily explores the "rubber reality" possibility (Keiper 2023, Ch. 16), a state where neither the characters nor the viewers are entirely certain whether they are in a dream or awake, as they easily and imperceptibly slip from one into the other. This concept directly parallels philosophical thought experiments, most notably Descartes' methodical doubt. Just as Descartes established the possibility that "everything I see is spurious" and that "body, shape, extension, movement and place are chimeras" (Descartes 1996, 16), Craven's film establishes the possibility that the boundaries between dreaming and waking consciousness are more permeable than it seems.

Craven's explicit philosophical intention becomes clear in his statement that "sleep is equated with a lack of knowledge of the truth" and that "to survive you must be awake, you must know what the truth is, and you must face it, and deal with it" (Craven, cited in Bukta 2018). This directly connects the film's central metaphor to fundamental epistemological concerns about the relationship between consciousness and knowledge.

The film also establishes a second, related but distinct philosophical possibility that connects to Plato's Cave allegory: ordinary consciousness (represented in the dream world) might represent a lower level of reality, and that awakening (represented in the vigil state) can provide access to higher truths. In *The Republic*, the prisoners in the cave must be "awakened" and "turned

around" to perceive reality rather than shadows: "when one was freed from his fetters and compelled to stand up suddenly and turn his head around and walk and to lift up his eyes to the light" (*Republic* 515c). While the Cartesian possibility concerns whether our perceptions can distinguish reality from illusion, the application of the Platonic allegory in the film hypothesises that different states of consciousness provide access to different levels of reality.

In Nancy's case, her enforced wakefulness, maintained through caffeine and sheer will, functions as a form of philosophical vigilance that reveals truths about her community's past actions and Krueger's continued threat. The adults around her remain metaphorically "asleep," refusing to acknowledge the truth about their vigilante killing of Krueger, just as Plato's cave prisoners resist enlightenment and "would kill anyone who tried to drag them out and lead them up" (*Republic* 517a).

These represent distinct epistemological challenges: the Cartesian possibility questions whether perception can reliably distinguish reality from illusion. In contrast, the Platonic possibility assumes truth exists but requires vigilant consciousness to access it. Unlike Cartesian scepticism's universal doubt, the film's Platonic framework remains optimistic: Nancy can "turn around" through wakefulness to perceive truths unavailable while asleep.

A Nightmare on Elm Street constructs its thought experimental scenario by systematically undermining the conditions that usually allow us to distinguish between dreaming and waking consciousness, while simultaneously creating circumstances where sustained vigilance reveals hidden truths. The film establishes both philosophical possibilities through Nancy's experience as she moves from ordinary consciousness to a state of "enforced" philosophical awakening.

The film constructs both scenarios simultaneously: escalating perceptual uncertainty (physical evidence from dreams, shared experiences with other characters, real consequences) establishes the Cartesian possibility, while Nancy's forced transition from passive sleep to active vigilance embodies the Platonic awakening. Her use of coffee, caffeine pills and willpower functions as a philosophical practice, revealing truths unavailable to the "sleeping" adults around her.

The film's thought-experimental structure culminates in Nancy's final confrontation with Krueger, where she must simultaneously navigate Cartesian uncertainty (the ambiguous ending questions whether any of her experience was "real") while maintaining Platonic vigilance (her survival depends on sustained conscious awareness rather than passive acceptance of appearances).

6.3.1 Cinematic techniques

Craven employs distinctively cinematic techniques to establish these philosophical possibilities through specific scenes that would be difficult to replicate through purely textual means. Each key sequence demonstrates how cinematic presentation can embody different aspects of the Cartesian and Platonic possibilities.

The most philosophically significant technique is what Muir calls Craven's "stunning transitions from reality to dream," where "characters glide from one plane of existence to the other without realizing it" (Muir 2004, 114). This technique directly supports the philosophical possibility that the boundaries between different states of consciousness are more fluid than commonly assumed. For example, when Nancy falls asleep in her English class, "we only realize [she is dreaming] when we see her dead friend standing in the doorway in a body bag, while a classmate's reading of *Hamlet* crawls down to a bone-chilling whisper" (Bukta 2018)

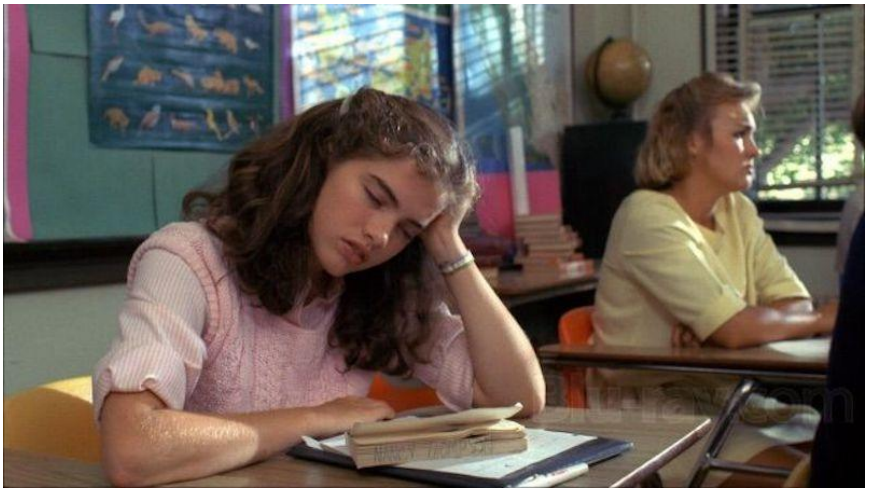


Figure 26: Still image - Nancy's school dream sequence, dreaming in Krueger's dreamworld.



Figure 27: Still image - Nancy's school dream sequence, dreaming in Krueger's dreamworld, where Tina's body is dragged.

This seamless transition establishes the possibility that we might unknowingly slip between different levels of reality without clear indicators.

The film's most direct embodiment of Cartesian scepticism occurs when Nancy falls asleep in her bathtub. What begins as an ordinary moment of relaxation becomes a terrifying intrusion when Freddy's razor-gloved hand emerges from between her legs in the bathroom.





Figure 28: Still images - Bathtub dream sequence: Nancy falls asleep in the bathtub and is dragged under it in the dreamworld/reality.

This moment literalises Descartes' concern that our most basic perceptions must be unreliable – if we cannot trust the most basic, private moment, what can we trust? The scene's

power depends entirely on cinematic presentation: viewers experience the same perceptual shock as Nancy, witnessing the impossible penetration of dream logic (being pulled into the bathtub water) into waking reality (where Nancy is just resting in the bathtub). Unlike a textual thought experiment that describes such an event, the film forces viewers to witness the actual breakdown of physical laws.

As Craven himself noted, "Since I started off as a writer, I tend to be verbal, so I love it when I can get myself to do something totally visual like the bathtub scene" (Craven 1988), recognising its cinematic power to provoke scepticism about our perception.

The film's use of physical objects that cross between dream and reality serves similar purposes. When Nancy pulls Krueger's hat from her dream into the physical world, the film establishes that "the logic of dreams seems to encroach more and more upon what the audience is meant to perceive as reality, and it becomes increasingly possible that the real world has also interposed upon the dream world" (Keiper 2023, Ch. 16).



Figure 29: Nancy holding Krueger's hat in the physical world after extracting it from her dream.

The film embodies the Platonic possibility through its presentation of adults who refuse to acknowledge the truth. Nancy's warnings about Krueger are repeatedly dismissed, embodying Plato's observation that cave prisoners "would kill anyone who tried to drag them out and lead them up" (*Republic* 517a). When Nancy's mother reveals the truth about burning Krueger alive, she immediately returns to denial: "He's dead, honey, because Mommy killed him." The adults' refusal to acknowledge Krueger's continued threat parallels the cave prisoners' resistance to enlightenment.

Craven's direction emphasises this through visual techniques: adults are often shown sleeping, drinking, or turning away from Nancy, while she remains alert and facing toward the camera. This visual strategy reinforces the film's central idea that wakefulness equals access to truth while sleep represents wilful ignorance.

Temporal inconsistencies provide additional evidence for the possibility of reality's instability. During the final sequence, Nancy

> calls her father and asks him to come break the door down in exactly twenty minutes. She then sets an elaborate array of booby traps around the house, has a heart-to-heart conversation with her mother Marge and goes to bed where she sets her watch alarm for ten minutes, indicating that it has only taken her ten minutes to do all that has happened since her conversation with her father" (Keiper 2023, Ch. 16).

These temporal impossibilities establish the possibility that what appears to be waking reality might be operating according to dream logic, supporting both the Cartesian concern about perceptual reliability and the need for Platonic vigilance to discern truth.

The film's conclusion returns to pure Cartesian scepticism by suggesting that Nancy's entire experience may have been a series of nested dreams. When Nancy appears to defeat Krueger by withdrawing her fear ("You're nothing. You're shit."), the apparent resolution dissolves into a new nightmare scenario. The final sequence, where Nancy enters a car that transforms into Krueger's realm, establishes the possibility of "multitiered dreams" (Keiper 2023, Ch. 16) that what we take to be waking reality might be nested layers of consciousness.







Figure 30: Still images - Nancy appears to have returned to normal reality, but then dream logic returns, suggesting Nancy's entire experience may have been nested dreams.

This technique is distinctively cinematic: the seamless visual transition from apparent reality to nightmare challenges viewers' perceptual confidence in ways that textual description cannot achieve.

6.3.2 Collaborative reasoning

We can find external support for this interpretation in Robert Englund's observation that he views "the entire film *Nightmare on Elm Street* as a precognitive nightmare of Nancy" (Englund, cited in Keiper 2023, Ch. 16). These scenes demonstrate how the Cartesian and Platonic possibilities reinforce each other in the film. Nancy's Platonic awakening (staying conscious to access truth) becomes necessary precisely because Cartesian scepticism proves valid (she cannot trust ordinary perceptual boundaries). The film suggests that philosophical vigilance is required not just to access hidden truths, but to navigate a reality where the normal indicators of waking consciousness are unreliable.

To understand *A Nightmare on Elm Street*'s philosophical content as a possibility-establishing thought experiment, it is necessary to examine how Craven's intentions align with the film's internal evidence. Craven's explicit statements, as we have seen, provide external evidence of philosophical intent regarding the relationship between consciousness and knowledge. His philosophical background supports the interpretation that these epistemic concerns are deliberate rather than accidental. Craven's personal and intellectual background reveals the deliberate philosophical foundations of the film. His training in psychology and philosophy proves important to this film's exploration of consciousness and knowledge. He traces the film's origins to his childhood experiences with the boundary between waking and sleeping consciousness:

> There was something about me that was always drawn to dreams. I had nightmares when I was a kid [...]. There was one germ that *A Nightmare on Elm Street* was based on – nightmares – and I was terrified to go back to sleep. In my child's naiveté I asked my mother to come to bed with me, and she said that's the one place I can't come with you. And I was totally awake, just like, "What!? What do you mean!?" She said she would be there when I woke up, but I would just have to be brave. I remembered that moment for a long, long time (Craven 2015).

More significantly, his philosophical and psychological research directly informed the film's epistemological framework:

> At the time I was reading about a level of consciousness in sleep where you are not really aware, yet are essentially fully awake; you need to go down a few levels or you need to lose your ego and break through [to a dream state]. That seemed like a perfect metaphor for a person trying to stay awake in order to face an overwhelming horrible truth that needed to be dealt with. As long as I scared the shit out of people every eight to ten minutes, I thought, let's try it! (Ibid).

However, the film's philosophical work operates through the collaborative model I have established: Craven architects the philosophical conditions through cinematic techniques, while viewers execute the thought experiment by recognising and exploring the possibilities the film establishes. This collaboration succeeds because Craven's explicit intentions align with the film's internal evidence: the seamless transitions between states of consciousness, the temporal inconsistencies, the physical objects that blur the line between dream and reality, and the ambiguous ending all support his stated epistemic concerns.

This collaboration requires viewers to actively construct mental models that accommodate impossible scenarios such as objects that blur the line between dream and reality, and temporal inconsistencies that violate normal causation. Unlike counterexamples, possibility-establishing experiments require viewers to expand their conceptual frameworks to include previously unconsidered scenarios. Viewers must also recognise these scenarios as testing universal claims about the nature of consciousness, reality, and knowledge rather than merely experiencing them as fantasy entertainment.

6.3.3 Cinematic advantages for possibility-establishing

A comparison between *A Nightmare on Elm Street* and *The Last House on the Left* reveals some differences between how possibility-establishing and counterexample thought experiments function in cinematic contexts.

Possibility-establishing thought experiments like *A Nightmare on Elm Street* appear particularly well-suited to cinema's distinctive capabilities. The film's establishment of "rubber reality" depends essentially on cinema's ability to manipulate visual and temporal continuity in ways that textual description cannot replicate. When dream objects appear in physical reality, viewers experience the perceptual breakdown directly rather than imagining it through linguistic description. This embodied experience of impossible scenarios may be more philosophically

powerful than textual presentation because it activates the quasi-observational processes more completely: we see the impossible rather than merely conceiving it.



Figure 31: Still image - Krueger emerging from the wall above Nancy's bed, demonstrating cinema's unique ability to visualise impossible scenarios.

Cinema's temporal manipulation gives it another advantage. The time inconsistencies and smooth transitions between dream and reality let us experience the uncertainty that the thought experiment explores. Viewers experience the same disorientation that Cartesian scepticism conceptually describes, but in a vivid way, rather than in an abstract manner, making the philosophical problem visceral rather than merely conceptual.

Counterexample thought experiments like *The Last House on the Left* may be more vulnerable to the objection that their particularity limits their philosophical value; not that it does not have it, but it may be reduced. By contrast, possibility-establishing experiments can achieve greater philosophical generality and thus have greater epistemic value. *A Nightmare on Elm Street*'s establishment of perceptual uncertainty applies to any epistemic situation, not just those involving supernatural threats. The film's philosophical work remains relevant even to viewers who reject its specific supernatural premises, as the possibilities it establishes operate at the level of consciousness and knowledge.

The different cognitive demands these experiments place on viewers also suggest different philosophical strengths. Counterexample experiments require viewers to recognise challenges to existing intuitions and potentially revise their moral or conceptual commitments. Possibility-

establishing experiments require viewers to expand their concepts and consider previously unconsidered scenarios. Both types of cinematic thought experiments are philosophically valuable, but they serve different functions.

This suggests that cinema might be especially good at establishing possibilities because of the medium's distinctive capabilities, such as visual manipulation, temporal control, and perceptual immediacy, which directly support the cognitive work these experiments need. While both types can succeed, possibility-establishing experiments may represent a more distinctive contribution to philosophy through cinematic means.

The film's continued cultural relevance as a reference point for philosophical discussions about consciousness and reality, similar to how Descartes' evil demon or Putnam's brain-in-a-vat scenario functions in academic philosophy, suggests its success in establishing these possibilities and demonstrates the collaborative success between Craven's cinematic architecture and viewers' philosophical engagement.

6.4 *Scream*: metahorror, hyperreality and media effects

Scream (1996) represents a complex case to analyse, demonstrating how a film can function simultaneously as multiple types of thought experiments while engaging in metahorror reflection. Unlike *The Last House on the Left*'s focused counterexample structure or *A Nightmare on Elm Street*'s possibility-establishing function, *Scream* serves many different purposes: it comments on the structures of horror film itself; it establishes a hyperreality possibility of the collapse between fiction and reality; and it serves as a counterexample to what we can call the media effects assumptions (that media is responsible for violent behaviour). The film engages in metahorror commentary, explicitly philosophising about the nature of horror cinema, genre conventions, and the relationship between fictional violence and audience consumption. *Scream*'s philosophical density effectively demonstrates how films can achieve a level of theoretical complexity that rivals traditional philosophical works while remaining accessible to a broad audience. I begin my analysis by examining the film's metahorror nature. Here I am are focusing on the first four *Scream* films of the franchise (1996, 1997, 2000, 2011) as they all follow the main character, Sidney, and progress in a coherent manner.

Scream follows Sidney Prescott (Neve Campbell), a high school student in the fictional town of Woodsboro, California, who becomes the target of a masked killer known as "Ghostface."



Figure 32: Still image - Ghosface mask.

The film opens with the brutal murder of Casey Becker (Drew Barrymore) and her boyfriend Steve Orth, establishing the killer's pattern of taunting victims with horror movie trivia before attacking them. As more students are murdered, Sidney realises she is being specifically targeted, with the killer making threatening phone calls that reference her mother's rape and murder one year earlier.

The opening scene with Casey shows something ironic about knowledge and survival. The killer tests her horror movie knowledge before killing her: her expertise becomes the tool of her own death. This sequence demonstrates that theoretical understanding of violence, i.e., knowing the "rules" of horror, provides no protection against actual violence and may even facilitate it. Casey's encyclopaedic horror knowledge becomes a ritual of participation rather than a tool of survival, suggesting that intellectual engagement with violent representation creates a false sense of mastery over real danger. The scene raises the question of whether consuming and analysing fictional violence actually prepares us for authentic threats or merely creates an illusion of preparedness that makes us more vulnerable. Randy, as the horror expert, has a similar fate in the next movie, showing how dominating horror tropes might not be good enough to survive.

The killings intensify around Sidney's social circle, including her boyfriend Billy Loomis (Skeet Ulrich), her best friend Tatum Riley (Rose McGowan), and horror film enthusiast Randy Meeks (Jamie Kennedy). News reporter Gale Weathers (Courteney Cox) arrives to cover the story, believing that Cotton Weary, the man convicted of killing Sidney's mother, is innocent. The film's climax occurs at a house party hosted by Stu Macher (Matthew Lillard), where Sidney discovers

that both Billy and Stu are the killers. Billy reveals his motivation: Sidney's mother had an affair with his father, which led to his mother abandoning the family. The killers plan to frame Sidney's father for the murders and kill Sidney as their final victim, but Sidney survives. The film spawned several sequels that continue Sidney's story while expanding the franchise's exploration of media, violence, and genre conventions.[18]

6.4.1 *Scream* as metahorror

Scream has been commonly considered a metahorror film. This agreement stems from the fact that *Scream* is aware of and exposes what the horror subgenre of slasher essentially is.

The slasher subgenre emerged in the 1970s and remained popular until the end of the 1980s, often following a "very simple formula," as Randy says. *Halloween* is sometimes considered one of the seminal slasher films – *Scream 1* seems to use this film as the template.

To understand the deconstruction that *Scream* makes of the slasher film genre, let us examine some of the standard features we expect to find in that subgenre.

According to Carol Clover's "list" of features we expect from slashers, the first one is the type of *killer* we find. The psychokiller who slashes his victims is usually a 'male in gender distress' (Clover 1993, 27), a male who has some kind of gender or sexual disturbance. As Clover points out, we can see such cases in *Psycho* (Hitchcock, 1960), where the killer, Norman Bates, is gender-confused or has two genders within him (his and his mother's). She also gives the example of *A Nightmare on Elm Street (Craven 1984),* where the killer, Freddy Krueger, is a child molester[19]. We could also point, for example, to *Friday the 13th*'s (Cunningham 1980) killer, the mother of Jason Vorhees, who seems to punish teen couples' sexual activities because her son, Jason Vorhees, (who becomes the killer in the sequels), drowned due to the lack of attention from his teen monitors who were more interested in sexual behaviour.

The *place* where the action happens is usually terrible. It can be a house, or a tunnel inhabited by terrible people or families (Clover 1993, 30). Norman Bates' or Michael Meyers (*Halloween*'s killer) are associated with creepy houses; Jason haunts a lake camp and its surrounding woods; Freddy taunts and kills children in a tunnel-like boiler room; the cannibalistic

[18] There are several sequels, including *Scream 2* (Craven, 1997), *Scream 3* (Craven, 2000), *Scream 4* (Craven, 2011), *Scream* (Bettinelli-Olpin and Gillet, 2022), *Scream VI* (Bettinelli-Olpin and Gillet, 2023) and an upcoming *Scream 7* (Williamson, 2026). Here we will only be focusing on Craven's films, not only for maintaining the same author (filmmaker), but because in the most recent sequels the focus shifts from Sidney's story.

[19] That Krueger is a child molester is not made explicit in Craven's original *Nightmare*.

family in *The Texas Chain saw Massacre* (Hooper 1974) live in a creepy house full of skeletons and cadavers in the middle of nowhere.

The kinds of *weapons* that are used, as the name suggests, are weapons that slash. Guns are not used (except for the victims). Freddy uses his glove with claws, Jason uses spears, machetes, and other weapons, Michael Myers usually uses a kitchen knife, Leatherface from *The Texas Chain Saw Massacre* uses a chainsaw, and so on.

The *victims* are usually teens, especially sexually active ones. Killers do not have any preference for victims of any specific gender; they kill boys and girls. According to Clover, boys die because they make mistakes, but girls die because they are female; as the males' death is 'nearly always swift', where he is dispatched fast, when girls die, their terror is registered in detail, as are their tortured bodies (Clover 1993, 34-35).

Slasher films also have a *final girl*, the girl who survives the killer by killing him herself, or by escaping and being rescued somehow. This girl is typically not sexually active and does not behave or present herself in the stereotypical way often associated with females (Ibid., 35-40). The final part of the film is focused on her struggle to survive.

Finally, slasher films are *shocking*: they often show graphic images and typically present gory content. As Clover says, "we see heads squashed and eyes popped out, faces flayed, limbs dismembered, eyes penetrated by needles in close-up and so on" (Ibid., 41).

Scream works with these slasher features and with audience expectations. That is the main reason why it is considered metahorror. *Scream* teaches us what to expect from a slasher film. According to Jancovich, "films of this subgenre are supposedly concerned with a process of terrorization in which a serial killer methodically stalks a group of teenagers who are killed off one by one" (Jancovich 2002, 5). This is precisely what happens in *Scream*. Let us examine some of the moments where Scream showcases slasher tropes and subverts our expectations for the subgenre.

In the first film (1996), Sidney's mother had been raped and murdered, her father is always travelling for work, and her boyfriend, Billy Loomis, is pressuring her to have sex. The man who was accused of her mother's murder, Cotton, claims to be innocent, something that the reporter Gale Weathers supports, even though for the wrong reasons (she wants to write a book based on Cotton's innocence and win a Pulitzer prize). As a spree of murders starts to happen again in Sidney's hometown, Woodsboro, she also starts to question if Cotton was her mother's murderer. She wants to believe that Cotton was her mother's rapist and killer, because there were rumours

that her mother was having sex outside the marriage, and that she slept willingly with Cotton, who was later framed for her murder.

The film begins with a girl preparing popcorn to watch a horror film, only to receive a call. When she answers the phone, she starts flirting with the voice on the other side, as they talk about their favourite horror films. Soon she realises that the voice is actually from a killer.



Figure 33: Still image - Casey answers the phone and talks to a stranger (Ghosface), asking horror film trivia.

He asks her if she wants to play a game and starts asking her questions about horror films, such as the name of the killer in the first *Friday the 13th*. The killer ends up killing her, starting the murder spree in the city, and raising the suspicion that Cotton is not the killer after all. This also becomes the killer's *modus operandi*: he calls the girls and taunts them, asking them about horror films, always ending up by killing them (except, of course, for Sidney, who always escapes).

We are introduced to several characters who take interest in these murders that soon become subject of a press circus: deputy Dewey, Sidney's best friend Tatum's brother, the reporter Gale, Sidney's friends, specially Randy, the horror geek who knows every slasher trope, and her boyfriend, Billy who is the first suspect as the killer.

Throughout the film, as in its sequels, various references to other horror films, particularly slasher films, are presented. We see that in the opening scene, as the first victim, Casey discusses and answers the killer's questions about slasher films. We also see teenagers watching *A Nightmare on Elm Street*, *Halloween*, or making various references to horror in their conversations. We are also faced with characters who know slasher films' tropes and "rules" to survive, especially Randy, such as the "rule" that dictates that sexual activity and bad behaviour mean getting killed. Because of such slasher conventions that follow a simple formula, Sidney tells the killer on the

phone that she dislikes horror films, mainly because they are so predictable and even insulting to girls who often make poor decisions, such as running upstairs instead of escaping the house.

In *Scream 2*, we also see numerous references to slasher films, but the attention now shifts to another topic related to Hollywood films in general. One of the ways by which it does that is by reflecting on the concept of sequel. In this film, Debbie Loomis (Sidney's boyfriend, Billy Loomis's mother, seeking revenge for her son's death) tries to make a sequel to the events in the first *Scream*. She does so by repeating the exact structure of the first "film", as she kills characters with the same names. Of course, Randy is aware that the driving motive behind the events taking place at Windsor University, where he and Sidney are now studying, is to make a sequel, as he also explains to Gale and Dewey the potential outcomes.

Scream 3 also reflects on Hollywood's strategies, specifically those related to trilogies. The new killer, Roman, who is also the filmmaker of *Stab 3* (the third instalment of the film based on Gale's book, which recounts the events in Woodsboro from the original Scream), is attempting to create a trilogy for both "reality" (*Scream*) and *Stab 3*. Here we see a tape filmed by Randy, who, aware that he would probably not survive the killer, tells us about trilogies and what happens in them. One of the significant scenes that happens in this film is in a scenario that reproduces the "real" Woodsboro: in this scenario we see what goes into making a film, even when we see, for example, doors that go into nowhere (when Sidney is running from Ghostface, she opens a door and almost falls because there is nothing there). This disallows "the suspension of belief that allows for cinema realism" (Jackson 2013, 17). This is also something that Craven had previously done in *Wes Craven's New Nightmare* (1994), since the film revolves around the production of *A Nightmare on Elm Street* that exposes "the making of" the film, and the process of writing by Craven himself.

In *Scream 4,* one of the themes is reboots. As *Stab* is in its seventh sequel and the audience is tired of sequels, a reboot is required to refresh the franchise. That is what the new killer, Sidney's cousin Jill, is trying to achieve: a reboot where she is the final girl. Here we see Craven heavily criticising reboots as films that are inferior to the originals since they rely on the other film's success, and that, in the long run, end up being forgotten.

In general, there are specific subjects that recur throughout the four films, reflecting on which may confer them the status of metahorror.

As Davinia notes, "each film also starts and ends with a reference to the social ritual of watching films and television" (Davinia 2006, 144). The film forces "the audience to question their

own pleasures and assumptions even as they watch" (Ibid.,140). So, *Scream* is also about the audience experience of watching films.

Although *Scream* shows us the slasher films' conventions and stereotypes, as their tendency to "kill" black characters or to show naked girls for no apparent reason, it also does not mesh with such conventions. Because the motive for the killings is to create a slasher film, the reality in Scream is not a slasher film, but, as Craven says, a thriller. When we see tropes being applied by the killers – using knives, using a mask, attacking teenagers – we also see slasher rules being different in "real" life, where losing virginity does not equal death anymore, where the killer is not always male, or where black characters do not always die, and so forth. Moreover, *Scream* questions genre conventions *in general*.

"Genres can be defined by various things: the characters, the theme, the stylistic patterns as music track, sound effects, lighting, or iconography (symbolic images)" (Bordwell et al. 2017, 330). However, as these are the conventional elements of a genre, for the audience, genre conventions shape our expectations about what we are likely to see and hear (Ibid., 329). These audience expectations are, for some theorists, the key element of genre.

According to Gregory Currie, genre enables works to have implicit content. By presenting genre elements, the audience knows what to expect from the film and is aware of the kind of plot and imagery it will have. Such "genre-based implications" are, thus, able to create certain expectations (Currie 2005, 45-46). Not only that, but these implications also help create "contrasting effects": viewers experience contrasting feelings regarding the film, where one can expect that *x* is going to happen and still hopes that y will happen. However, if our reasons for expecting x are based on the film genre, this expectation is not present within the fictional story, i.e., the characters are unaware of their fate (Ibid., 46). Perhaps I could add that this is the reason why we hope something different from what happens in genre films occurs in that film, because characters act unknowingly of their genre-fates, we are led to sympathise with them in such a way that we hope they can survive or overcome certain situations. Moreover, perhaps the "contrasting effects" are the reason we keep watching genre films, because although we know the characters' fates, we still feel for them.

Furthermore, Currie states that a genre is a "genre-for-a-community". Thus, knowing what that genre is is a contingent and empirical question (Ibid., 48). In the same vein, Andrew Tudor argues that "[g]enre is what we collectively believe it to be" (Tudor 2012, 6). Steve Neale also argues that genres consist "of specific systems of expectation and hypothesis that spectators bring

with them to the cinema and that interact with films themselves during the course of the viewing process" (Neale 2012, 179). Both agree that defining genre in terms of essential features seems to be futile, since "most elaborate definitions always seem to throw up exceptions" (Ibid.,189), but still put the weight of genre, as Currie does, in audience expectations. For Mark Jancovich, such a view can also be problematic: in reality, a "policing" of the constitution of genres happens outside of academia, i.e., even the audiences "struggle over the definitions of genres" (Jancovich 2000, 23). Thus, as with other genres, "[t]here is no simple, single 'collective belief' as to what constitutes the horror genre, even among self-identified fans of the genre" (Ibid., 25).

It is commonly thought that horror aims at provoking a particular effect in viewers: in particular, horror films aim to provoke horror: "the horror genre is most recognizable by the emotional effect it tries to arouse. The horror film aims [...] in short, to horrify" (Bordwell et al. 2017, 340). For Noël Carroll, "raising various preordained emotions in spectators is the *sine qua non* of certain film genres", which is the case, for him, of horror film (Carroll 1999b, 35). Also, for Cynthia Freeland, "horror films are designed to prompt *emotions* of fear, sympathy, repulsion, dread, anxiety, or disgust" (Freeland 2000, 3). Wes Craven himself articulated his view on the nature of horror: "Horror films in general are about the terror event entering adulthood because you think adulthood is actually evil" (Craven 2005). This suggests that, for him, horror's emotional impact serves a broader existential function, examining the psychological transition from innocence to experience.

We can generally disagree with an account of the horror genre based on audience expectations of feeling scared, and it could be said that horror fans disagree on what emotions they expect certain films to cause. However, Currie's idea of contrasting feelings seems to be adequate for horror viewers. As certain genre films follow some kind of formula, viewers expect that certain events are going to happen, characters are going to behave in specific ways, kinds of weapons are going to be used, and "rules" are going to be followed.

According to Currie, the expectations we have for films are unknown to the characters. This difference in epistemic position between characters and viewers is one of the reasons why Carroll argues against the identification theories of the paradox of fiction, as we saw earlier. For Carroll, there is an emotional asymmetry between characters in the fiction and viewers. This asymmetry can also be attributed to the differing epistemic positions between viewers and characters. As characters "really" suffer from the film's events, viewers are in a privileged epistemic

position of knowing that they are watching a fiction and knowing the genre conventions that could lead to specific fates.

Scream is aware of this audience's epistemic position and expectations for a slasher film and plays with the idea of horror as a formulaic genre. As Jancovich says, "*Scream* (1996) both draws upon and reproduces in its supposed self-referential play with the 'rules' of the horror film. Not only does Randy Meeks suggest that there are certain inevitabilities within plot and character, but Sidney Prescott even describes horror films as 'all the same'" (Jancovich 2002, 12).

In *Scream*, the contrasting effects of a typical slasher film and the typical epistemic positions of viewers and characters are confused to the point where the characters know what to expect, while viewers do not. In the end, characters and viewers' expectations are simultaneously met and betrayed. *Scream* appears to reflect on the nature of genre and audience expectations, and to play with and mix them. However, it does not do that solely by stating it in dialogues and particular scenes; it shows us that by doing it, i.e., by doing that to the audience.

There is another subject about which *Scream* reflects, turning it into a self-reflective film, or a meta-horror film: the relationship between horror films and real violence. In this context, Scream heavily criticises the news media's obsession with and representations of real violence. This is also what I will argue to concede *Scream* the thought-experimental status.

Scream's metahorror elements, like the characters' explicit knowledge of genre conventions and the film's self-reflexive commentary, function primarily as cinematic techniques that enable both the media effects counterexample and the hyperreality investigation I examine next.

6.4.2 *Scream* and the hyperreality possibility

Related to this franchise's metahorror nature is its exploration of blurring the lines between fiction and reality. This relates to the concept of hyperreality, which is why I argue that Scream establishes that possibility.

The concept of *hyperreality*, most notably explored by Jean Baudrillard, refers to a state of affairs where there is no longer a reality corresponding to representations. In this kind of state, some instances attempt to be fictional (such as Disneyland) in order to conceal the fact that there is no longer a reality to distinguish from fiction. In this case, there is no longer anything "real", but only instances of hyperreal.

This happens through a process of *simulation*: whereas "feigning or dissimulating leaves the reality principle intact", where there is a clear difference between the real and the fictional, "simulation threatens the difference between 'true' and 'false', between 'real' and imaginary" (Baudrillard 1998, 168).

In this case, images would go from reflections of reality to perversions of reality, then to masks of the absence of a fundamental reality, and finally, to completely unrelated to any reality, thus turning into their "own pure simulacrum" (Ibid., 170).

This is enabled by a society replete with media representations. As "real news" are 'too often subordinated to image manipulation', events like wars 'become in some way just 'dramatized media events' which take place on TV'" (Butler 2002, 111). In this case, TV just spreads 'fictionalized information' (Ibid., 111). According to these ideas "we live in a *society of the image*, primarily concerned with the production and consumption of mere 'simulacra'" (Ibid., 112). A case of this "fictionalised information" and "dramatized media events" was O. J. Simpson's trial broadcasted on TV, where the acts seemed to be played or acted (Ibid., 115). This would mean that the trial was a simulacrum or had become an instance of hyperreality: there was no longer a "real" referent to the "fictional" trial we were watching.

Scream establishes the possibility that this theoretical condition could become experientially real through the following demonstration:

1) Media saturation could theoretically collapse stable reality/representation boundaries through progressive simulation;
2) Characters could lose the reliable capacity to distinguish between fictional and real domains;
3) Representational layers could multiply to create infinite regression without a foundational reality;
4) Individual identities could become hyperreal simulations of fictional archetypes rather than authentic selves;
5) Therefore, hyperreality represents a genuine possibility rather than a mere theoretical abstraction.

6.4.2.1 The progressive collapse of reality: Scream (1996)

The first and most noticeable relationship between *Scream* and the concept of hyperreality is evident in the first film, particularly with the character of Gale Weathers. She is the news reporter

desperate to show everything that is happening in the city of Woodsboro on TV, and to show that Cotton is innocent – even, it appears, if that is not the case. All because she wants to write Cotton's story and win a Pulitzer Prize. However, the argument for *Scream* as a thought experiment on the possibility of hyperreality relies on other subtler and more significant aspects.

In her pursuit to know and record everything happening with the teenagers that are being killed one by one, Gale installs a camera in the house where they are having a party and watching *Halloween*. At times, we see these images that the camera is taking of the real events happening in the living room where the teenagers are. As the killer is in the house murdering again, and as Randy tells Jamie (the actor playing Laurie, *Halloween*'s final girl) to look behind her while the killer is behind him, we are watching layers upon layers of images: we see Gale's camera's images of Randy's image of him watching a fictional image (*Halloween*).

As Tietchen notes, "the representation of the event (personified by Weathers) becomes subsumed within the event itself, until the real and its simulation collapse inseparably together" (Tietchen 1998, 104). *Scream* works similarly. The scene that opens this article is an example of what I am arguing. Randy starts telling Jamie on the TV, "Watch out, Jamie, you know he's around. There he is. I told you. I told you he's right around the corner". At the same time, Ghostface, *Scream*'s killer, appears right around the corner to kill Randy, and at the same time, the *Halloween* creepy music starts. Then, as he screams to Jamie, "Look behind you, turn around", little does he know that the killer is behind him. And then, in the TV news van, the cameraman starts yelling to Randy, "Behind you, kid!" We are faced with layers upon layers of fiction and the experience of watching horror events.



Figure 34: Still image - Randy tells Jamie to look behind her, while a similar situation depicted in Halloween happens with him.

6.4.2.2 The progressive collapse of reality: Scream 2 (1997)

The second *Scream* instalment begins with teenagers attending a movie screening to watch Stab, the film adaptation of Gale's book about the Woodsboro murders. Here, we watch an image that is an adaptation of a book based on "real" events. Teenagers run through the cinema theatre, masked with promotional fake Ghostface costumes and fake knives, identical to the original, albeit fake. As this happens, the real Ghostface is among them, with, of course, the same costume. In this hyperreal environment, killing becomes remarkably easy. So, that is what happens: Ghostface murders Phil, the boy who convinced his girlfriend, Maureen, to watch *Stab*. After that, he sits next to Phil's girlfriend. Not noticing that it is not Phil next to her, because he his masked, she talks to him normally, until she realises something is wrong and it is too late, as she gets stabbed. She is stabbed multiple times as she walks to the stage, bleeding and dying: the audience just watches this as they would be watching a film.



Figure 35: Still image - Maureen dies in front of the screen during the screening of *Stab*.

As Davinia notes, "Maureen dies because the *Stab* audience can't separate fiction from reality" (Davinia 2006, 146). The audience cannot separate fiction from reality because it seems that there is no longer any distinction between the two; fiction has taken over.

Another important event is Cotton's negotiation with Sidney for a TV interview, where the story of his innocence is told. Throughout the film, Sidney resists this idea, which appears to be her resistance to falling into the media circus that has turned her real life into a horror film. But by the end of the film, Cotton helps Sidney to survive one of the killers (Debbie Loomis) with the condition that she would grant him that interview, which she reluctantly does. When everything is over, and the killers are dead, the media is outside the college theatre where that happened, waiting

to record and report everything they can. When they go to Sidney, she points to Cotton, saying that he was the hero. For the media that is reporting on the event, that is the real story. Cotton triumphs with the hero narrative, as the following film shows.

6.4.2.3 The progressive collapse of reality: Scream 3 (2000)

In the third film, we learn that Cotton hosts his sensationalist TV show, where he discusses the increasing violence in society. We also learn that he is taking part in the production of *Stab 3*, which does not happen because he is killed in the first scenes. Since there is no *Scream 3* yet, we can deduce that there was a Stab 2, which served as the fictional version of the real events of *Scream 2*. So now, *Stab 3* has no real ground to be based on.

We learn that Dewey is also involved in *Stab 3* as a "reality-checker," or someone who will ensure the similarity of scenarios to the original Woodsboro. In *Stab 3*, Gale will be the killer. We see that the script is being written in the meantime.

With Cotton's death, Gale is immediately drawn to Hollywood to investigate what is happening, where she finds Dewey. As the police start to investigate, Detective Mark Kincaid asks Dewey to call Sidney to go there and help. Reluctant, but knowing that she cannot hide, as it seems that the killer has already found her, she goes there.

With Cotton's death, the studios feel pressured not to make a *Stab 3*, as if the film were responsible for this death. However, the script continues to be written, and we meet the actors who impersonate the real trio (Sidney, Gale, and Dewey). Actors die as the script appears to be written into that fate right before their deaths. The fictional script seems to create reality.

One important detail is the scene where Sidney escapes Ghostface in the Woodsboro scenario: she is, once again, running from her killer in her "home" or, at least, the representation of it. By the end of the film, it becomes clear that the killer is actually the filmmaker, Roman. As Sidney is running from her killer in reality, she is also running from the creator of a fictional version of her real killer: the Woodsboro scenario is now hyperreal, with the referent being itself.



Figure 36: Still images - Sidney enters a replica of her home, which serves as the scenario for *Stab 3*.

As Roman confronts Sidney by the end of the film, he reveals to her that he was his illegitimate brother, who was rejected by Maureen, as he was the result of rape she had suffered for not agreeing to sleep with producers to get film parts. He also reveals that he showed Billy videos of Maureen with his father that caused his mother to abandon his father, which finally led Billy to kill her.

Roman turned Sidney's life into a literal horror franchise. His motive is, thus, to create a true-story-based fictional film in which Sidney dies. Since *Stab 3* has no referent in reality, Roman starts to try to create referents for fictions, where there are none. There is no reality anymore, only a fiction desperately trying to grasp a reality.

6.4.2.4 The progressive collapse of reality: Scream 4 (2011)

With the fourth film, reality seems to have finally disappeared, and more than ever, there is an almost absolute confusion between reality and simulation. The film begins with this state of confusion:

> Two teen girls sit on a couch discussing scary movies; the phone rings, it's the killer's voice on the other end, and after an elaborate kill sequence, both girls end up dead. Then we cut to a new set of girls, who were apparently watching the first set in a scary movie. This time, one of the girls complains about 'the whole self-aware, postmodern metashit' and how it had been done before. Her friend takes out a knife and stabs her, asking, 'Did that surprise you?' Cut to yet another set of girls, who were in turn watching the first two sets as part of a movie – yet another metacommentary on horror – and then those two girls are butchered. It *could have gone on forever*, but finally these last two girls are revealed to be part of the diegetic reality of *Scream 4* [my emphasis] (Davinia 2006, 144).







Figure 36: Still images - A film within a film within a film.

This scene could have been, indeed, an eternal regression; at least, the effect it has on viewers' minds is that such an eternal regression of simulation within simulation is possible; in other words, that hyperreality is possible, with a sense of a loss of reality.

When the action starts to happen in *Scream 4*, we meet Jill, Sidney's cousin. The events and characters here seem to be an imitation (or reboot) of the original *Scream*. At this point, there is no longer an original Woodsboro. After years of TV coverage of the city, of fictions made about events that happened in it, and with events so similar happening again, it seems that Woodsboro succumbed to a fictional version of itself, a hyperreal Woodsboro; "it has come to reflect its own simulation" (Jackson 2013, 18). According to Jackson, Sidney "returns to Woodsboro only to find that it, and everyone in it, has become a simulation", pointing to Baudrillard's concept of hyperreality (Ibid.,12). This appears to be an accurate account of Scream 4 and the Scream franchise in general.

We learn that Jill is now the killer, and her reason to kill everyone, even her mother, is to become her version of *Scream*'s (or *Stab*'s) final girl. As her goal is to annihilate Sidney, it is as if she wants to turn into a hyperreal Sidney.

Then, in the hospital, when Jill finds that Sidney survived, she goes to her room trying to kill her again. As this happens, and Sidney finally kills Jill, the media outside reports the news: Jill is the new hero of Woodsboro. The news is fulfilling a fiction that Jill created, and even though we, as viewers, know what is really happening in the hospital room, this leaves us with an uncomfortable feeling that this probably happens in our real TV news: villains passing as heroes in a reality they created or distorted, and media that propagates it.

As the series progresses, we notice a process of simulation, or dilution of reality, through the four films, into what seems to be a hopeless hyperreality. As Jackson suggests, the "series stages a crucial demonstration of the evolution of the reality representation relation in an increasingly mediatized society, interrogating the ramifications of these trends for our understanding of originality and authenticity" (Jackson 2013, 14). Davinia also notes that *Scream* points "to the inability of culture to find a separate Truth ('with-a-capital-T') apart from the narratives that we invite inside our homes on a regular basis" (Davinia 2006, 144).

6.4.3 *Scream* as a counterexample to media effects assumptions

Scream can function as a counterexample to media effects assumptions that exposure to violent media directly causes violent behaviour. Against claims that horror films "undermine the reactive attitudes that are responsible for moral agents being disposed to respond sympathetically to instances of victimization" (Woodcock 2013, 309), Craven's film demonstrates how media effects

assumptions hide deeper psychological and social roots of aggression while providing convenient scapegoats for those seeking to rationalise violent acts. The film shows that the causal relationship between media and violence flows in the opposite direction from what the media effects assumption would claim.

One of the main arguments for the immorality of horror films centres on what Woodcock terms the "Argument from Reactive Attitudes" (ARA), mostly defended by Di Muzio (2006). This argument posits that horror films are morally problematic because they threaten to undermine the psychological capacities that enable appropriate moral responses to suffering, specifically our sympathetic dispositions. The ARA is composed of two main premises:

1) "It is prima facie morally wrong to view, or to facilitate viewing, those features of a work of art or entertainment that encourage the corruption of reactive attitudes that are necessary for human agents to develop and maintain a well-functioning moral psychology" (Ibid., 310). This argument is based on recognising that human moral agency depends on our sympathetic dispositions, which are not automatic but require cultivation and can be damaged over time (Ibid., 311).
2) "Taking pleasure in murder, torture, dismemberment and other acts of horror violence can threaten the proper functioning of our sympathetic attitudes" (Ibid., 313).

According to this perspective, horror films are perceived as having a corrupting or damaging influence on those who hold these dispositions. Di Muzio specifically targets slasher films as they are "devoted primarily or solely to representations of violence or death" (Di Muzio 2006, 281).

The media effects argument defending this position follows this reasoning:

1) Human moral agency depends on sympathetic dispositions that enable appropriate responses to suffering and victimisation;
2) These sympathetic attitudes require cultivation and can be systematically corrupted through repeated exposure to representations of violence;
3) Horror films encourage viewers to take pleasure in violence, torture, murder and other forms of suffering;
4) This pleasure-taking systematically undermines our capacity for moral sympathy by normalising violence and desensitising viewers to suffering;
5) Therefore, horror consumption corrupts viewers into morally diminished agents who lack appropriate reactive attitudes toward real violence.

This position finds support in broader media effects research and concerns about desensitisation, suggesting that repeated exposure to fictional violence creates psychological changes that transfer to real world moral responses.

Scream presents a counterexample to this reasoning:

1) The film presents characters who are extensive horror film consumers yet retain strong moral sympathies and appropriate reactive attitudes;
2) These horror-literate characters respond to real violence with genuine shock, revulsion, and moral concern rather than desensitised indifference;
3) Their knowledge of horror conventions actually enhances their moral awareness by providing frameworks for recognising and responding to danger;
4) The film demonstrates that actual killers use horror films as convenient post-hoc rationalisations rather than genuine causal motivations for violence;
5) Therefore, horror consumption does not corrupt moral psychology but may actually enhance moral awareness through critical engagement with representations of violence.

An important aspect of ARA is what Woodcock calls a "burden of proof reversal", where rather than requiring conclusive empirical evidence that horror films cause this kind of psychological damage, the argument shifts the burden to defenders of horror consumption. As he explains, "evidence should be provided to justify the risk one is taking to one's moral psychology against a default presumption that one ought to pursue morally innocent forms of amusement" (Woodcock 2013, 313). One of the main problems here is that it is difficult to measure these types of subtle psychological effects. As Stoner notes, "Slight damage to the appropriate functioning of a subject's reactive attitudes relative to a counterfactual no-exposure doppelganger is, to understate the case, difficult to measure. Empirical research may never decisively confirm or falsify the ARA's second premise" (Stoner 2020, 6). Thus, in the absence of conclusive empirical evidence, ARA proponents may argue that "the link between actively taking pleasure in sadistic violence and undermining one's moral capacities is sufficiently plausible [...] for it to be morally appropriate to adopt a precautionary attitude toward the choice to view such violence as entertainment when other options are available" (Woodcock 2013, 313).

Thus, another concern of ARA is about desensitisation. Rather than claiming that horror films directly cause viewers to commit violent acts, the argument focuses on psychological effects. As Stoner states, "Di Muzio and Woodcock are not committed to the claim that gorefests turn gore fans violent. They are only committed to the claim that gorefests make gore fans at least a little bit

less compassionate than they otherwise would have been" (Stoner 2020, 6). In this case, someone might not commit violent acts per se but still experience reduced emotional responsiveness to others' suffering.

Scream's counterexample first establishes the possibility of media effects as plausible: characters and media coverage initially suggest that horror films might be influencing the killings. Second, it provides increasingly detailed evidence that contradicts this surface explanation, revealing deeper psychological and social causation. Third, the film demonstrates that assumptions about media effects are false.

Billy Loomis exemplifies this progression clearly. His motivation for the Woodsboro murders stems from family trauma: Sidney's mother had an affair with his father, which led to his mother abandoning the family. When Sidney tells Billy, "This is life, this isn't a movie", he replies by saying, "It's all one great movie". Billy's response demonstrates how he uses horror film narratives not as causes but as aesthetic justifications for the violence that is actually rooted in psychological wounds.

The revelation that Billy and Stu work as collaborative killers exposes the social nature of hyperreality construction. Rather than individual psychopathy, their partnership demonstrates how shared fictional realities require collective participation. Their coordination reveals that turning "reality into fiction" demands collaborative agreement about which narrative framework will govern actual events. The killers function as co-directors of a shared delusion, forcing other characters to become unwilling participants in their fictional reality.

This pattern extends throughout the franchise. For example, in *Scream 2*, Mickey wants to use the existence of horror films in his defence, revealing what Craven called the "media circus" that turns real violence into entertainment while deflecting attention from genuine causes. Mickey's strategy reveals the film's central insight: media effects claims function as displacement mechanisms that allow violent actors to avoid confronting the authentic sources of their aggression.

The film shows that blaming the media for violence creates a shortcut that stops us from understanding what really causes it. As Stoner observes, "careful attention to statistical evidence shows no relationship (or even, possibly, a negative relationship) between violent media and violent crime" (Stoner 2020, 6). *Scream* dramatises this disconnect by showing how media effects assumptions impede recognition of actual causes.

Roman's revelation in *Scream 3* provides the clearest demonstration of this. Roman reveals that he was Sidney's illegitimate half-brother, rejected by their mother, Maureen, who was the

result of rape she had suffered for not agreeing to sleep with producers to get film parts. Roman also reveals that he showed Billy videos of Maureen with his father, which caused Billy's mother to abandon his father and ultimately led Billy to kill Maureen. Roman's orchestration of the entire film's violence stems from his rejection and trauma, not from horror film consumption. Significantly, as a filmmaker himself, Roman understands media manipulation and uses horror film conventions as tools for expressing pre-existing vengeful impulses rather than being shaped by them.

Scream's franchise provides a systematic counterexample iteration: each sequel reveals additional cases where media effects explanations prove inadequate while psychological and social factors provide consistent explanatory power. The counterexample does not merely show that media effects claims are false; it reveals them to be systematically misleading in ways that prevent recognition of violence's real sources. The film shows that the burden of proof for media effects claims cannot be met, not because the evidence is inconclusive, but because the causal relationship might flow in the opposite direction from what media effects theorists assume. Maybe, violent individuals do not become violent through media exposure, but rather they selectively interpret and deploy media narratives to serve their interests.

As we saw earlier, ARA presents a more subtle challenge to horror films. Rather than arguing that horror films cause violent behaviour, ARA contends that consuming violent entertainment desensitises viewers' emotional responses to suffering, potentially undermining the sympathetic dispositions necessary for moral agency. *Scream* provides counterevidence to these desensitisation claims through Randy's character arc. Randy represents the archetypal horror film enthusiast: he demonstrates encyclopaedic knowledge of genre conventions, shows apparent desensitisation to fictional violence by clinically analysing horror films, and treats movie violence as material for intellectual analysis. If ARA's desensitisation thesis were correct, Randy should exhibit reduced sympathetic responses to real suffering.

However, the film demonstrates precisely the opposite pattern. When confronted with actual violence against his friends, Randy displays intense emotional responses. His reaction to Tatum's death, his protective instincts toward Sidney, and his horror at discovering the true extent of the Woodsboro murders reveal that his apparent desensitisation only operates within fictional films. Randy's expertise in horror films does not diminish his capacity for appropriate moral responses to authentic suffering; instead, his genre knowledge helps him identify real threats and protect potential victims. This suggests a distinction that ARA fails to acknowledge: desensitisation

to fictional violence does not necessarily transfer to real-world emotional responses. The capacity for contextual discrimination challenges ARA's assumption that desensitisation effects generalise across fictional and real domains.

Furthermore, even if we were to accept that horror film consumption reduces sympathetic responses in some contexts, ARA requires a separate argument to establish that such a reduction is necessarily morally problematic. The argument assumes that any diminishment of sympathetic dispositions represents moral degradation. However, we should also consider essential social roles that explicitly require lessened sympathetic responses, such as medical professionals, emergency responders, or police officers. These professions demonstrate that reduced sympathetic responses can serve legitimate moral purposes rather than indicating moral corruption. Therefore, the moral distinction should be made not between the presence or absence of sympathetic responses per se, but between responses that are appropriate or inappropriate in context and function. A surgeon's emotional detachment during surgery serves patient welfare; a horror fan's reduced visceral response to fictional violence serves no harm while potentially providing benefits, such as genre appreciation or intellectual engagement with moral themes.

Scream represents a more complex case within our thought experiment taxonomy than the previous films examined. While *The Last House on the Left* functions primarily as a counterexample to revenge intuitions and *A Nightmare on Elm Street* serves as a possibility-establishing experiment for epistemic scenarios, *Scream* works as different types of thought experiments simultaneously. The franchise primarily establishes the possibility of hyperreality, demonstrating how media saturation might collapse distinctions between representation and reality. However, it also functions as a counterexample to media effects assumptions, challenging the widespread belief that fictional violence directly causes real violence. By extension, the film's counterexample to media effects assumptions confirms alternative explanations that locate the roots of violence in psychological factors rather than in the consumption of horror films. Thus, the film can also be seen as a thought experiment that confirms a theory.

6.4.5 Cinematic techniques in *Scream*'s thought experiments

Having analysed *Scream*'s functions as metahorror commentary, hyperreality possibility, and media effects counterexample, I can now examine the specific cinematic techniques that enable these philosophical engagements.

Scream employs distinctively cinematic techniques to create its philosophical commentary. As Craven stated about his approach to filmmaking, he aimed to give audiences "an experiential feeling" by pushing "the visuals to an extreme that kind of makes you feel like you're in that kind of outer edge of consciousness, into some kind of deeper subconscious experience of an event" (Craven, 1999). This experiential quality is fundamental to how *Scream* functions as a philosophical thought experiment about genre conventions and media effects.

The film's most significant technique involves its systematic use of layered visual composition to demonstrate the collapse between representation and reality. Two different scenes showcase exactly this. One of them is during Randy's exposition of slasher "rules"



Figure 37: Still image - Randy explains slasher rules.

Another important sequence occurs when Randy watches Halloween while being filmed by Gale's camera, creating a literal visualisation of Baudrillard's simulation stages. The editing employs a complex framing where Randy appears simultaneously in three different representational contexts – as a "real" person in *Scream*, as an image captured by Gale's surveillance camera, and as a viewer consuming Jamie Lee Curtis's performance in *Halloween*.

Craven's editing choices during this sequence do important philosophical work. Instead of conventional shot-reverse-shot patterns that maintain clear spatial boundaries, we encounter rapid cross-cutting. As we watch, this technique confuses our sense of which representational level we are observing.

This creates what Davies calls "critical commentary", where "visual elements provide a critical commentary on what is being said" (Davies 2019, 143). The editing does not merely

illustrate the rules: it demonstrates their simultaneous validity and inadequacy by showing characters both following and subverting them in real time. When Randy shouts "behind you!" to Jamie Lee Curtis while Ghostface approaches from behind, the editing creates temporal synchronisation between fictional and "real" events, making the distinction between them phenomenologically harder to maintain. As Randy warns the fictional character about danger she cannot perceive, viewers experience the same epistemic limitations Randy faces: we can see the approaching threat he cannot, yet we are powerless to warn him, just as he is powerless to help the fictional Jamie.

The sound design contributes essential content through its treatment of the characteristic distorted voice of Ghostface.



Figure 38: Still image - Billy shows Sidney how he has been using a voice changer when he calls her.

Unlike written descriptions of threatening phone calls, the film employs voice modulation technology not merely for atmosphere, but as a concrete demonstration of how media can create artificial personae that become indistinguishable from authentic identities. Billy's revelation of the voice changer technology represents more than mere deception; it demonstrates the complete dissolution of authentic identity in a mediatised environment. The device enables anyone to become anyone else through technological mediation, eliminating stable markers of authentic recognition. When Billy shows Sidney how easily he has been impersonating the killer's voice, the film reveals that identity itself has become a performative act. This moment suggests that in hyperreal conditions, the question is not "who is speaking?" but rather "which performance of identity is currently being deployed?" The voice changer literalises Baudrillard's claim that simulation threatens the distinction between authentic and artificial identity, making recognition itself impossible.

As Craven noted, his goal was always to create work that would "challenge, manipulate, and entertain" audiences through sensory engagement (Craven 2005). When Maureen in *Scream 2* sits next to the killer without recognising the substitution, the audience experiences the same perceptual limitations she faces. The identical costumes, combined with the voice distortion technology, create conditions where identity becomes purely performative rather than essential. This shows a cinematic way of exploring whether authentic identity might be impossible in a media-saturated world.

The franchise's framing techniques, which blur the boundaries between fiction and reality, develop across the sequels, similar to the techniques Craven used in A *Nightmare on Elm Street*, escalating to include film-within-film, dream-within-dream, and dream-within-film sequences. The set design in *Scream 3* gives us another cinematic tool. When Sidney runs through a copy of her childhood home, we see doors that open onto space and walls that are just facades. The film presents these impossible spaces without explanation, allowing viewers to discover the illusion of the environment alongside Sidney. This technique transforms the theoretical claim that reality might be constructed into a spatial experience.

Scream 4's opening sequence demonstrates cinema's unique capacity for infinite regression in ways that textual thought experiments cannot replicate. The film-within-a-film-within-a-film structure creates a kind of phenomenological vertigo, letting viewers feel the disorientation that Baudrillard describes. Each revelation that we have been watching a movie rather than reality forces viewers to reconstruct their understanding of what constitutes the "base level" of reality. The editing employs identical framing, lighting, and performance styles across all levels, making the transitions between levels imperceptible until they are explicitly revealed. This technique generates uncertainty about the nature of representation that purely linguistic descriptions cannot achieve.

These cinematic techniques succeed because viewers cannot easily dismiss the possibility of living in a simulated environment when they have directly experienced the impossibility of distinguishing between representational levels. The film's presentation provokes viewers to carry the experience of uncertainty about representation into their lives, making Baudrillard's theoretical claims feel like lived possibilities rather than abstract concepts.

6.4.5 Collaborative Reasoning in *Scream*

To understand how *Scream* operates across its multiple philosophical functions – as metahorror commentary, hyperreality possibility, and media effects counterexample – I must examine the relationship between Craven's film architecture and the viewers' engagement. Craven stated in an interview: "My intention is to make people aware of the media. By hypnotizing or involving them at one point and then distancing them at another, I can make them aware of the various states they're in while watching a movie, states they might not ordinarily recognize" (Craven 1990). This strategic manipulation of psychical distance explicitly requires viewers to perform metacognitive work: recognising their own states of consciousness while viewing.

From an authorial perspective, Craven's works demonstrate systematic concern with the relationship between fiction and reality. As Tietchen states,

> The flux between 'the real' and 'the representation of the real' is firmly fixed within Craven's intentions. [...] Viewers familiar with his *Nightmare on Elm Street* series know too well the nebulous boundaries of Craven's diegesis, the seamless separations (established through editing) that defy our traditional conceptions of reality and allow murderers like Freddy Krueger to posit their killings simultaneously in real and representational space" (Tietchen 1998, 130).

This was fully developed in *New Nightmare*, but reached its peak with *Scream*, created in collaboration with Kevin Williamson's script, who was also concerned about the escalation of violence broadcast on television. About the metahorror character of *Scream*, where teenagers explicitly know horror film rules, Craven explained the deeper philosophical work the film requires from viewers:

> Beneath the surface of that, because that's kind of the obvious look, there's a very interesting set of study of the kids' perception of what violence is and how do you deal with it. How it switches from their film idea. Because they are such fans of these films and find them entertaining, they have to change everything to account for it when it actually happens to them in real life. So, it is a very interesting sort of film within a film. [...] *Scream* is very much about the whole world of kids who watch these films and how it affects their idea of reality" (Craven 1997).

This collaboration between Craven and the viewers entails that, first, viewers must recognise the layers of simulation the film presents and track which representational level they are observing at any given moment. Second, viewers must resist the natural tendency to dismiss hyperreality as merely fictional and instead engage with it as a possibility. Third, they must navigate the emotional engagement the film demands while maintaining analytical distance to process its philosophical implications about media, reality, and violence.

The success of this collaboration is evident in the film's continued cultural relevance as a reference point for discussions about media effects, reality, and representation, much like traditional philosophical thought experiments in academic discourse.

6.5 *Wes Craven's New Nightmare*: an archetypal media theory

Wes Craven's New Nightmare (1994) represents perhaps the most philosophically ambitious entry in Craven's filmography, functioning simultaneously as a horror film, a meta-commentary, and an investigation into the nature of reality, representation, and authorial responsibility. The film follows actress Heather Langenkamp, playing a fictionalised version of herself, as she navigates increasingly disturbing events that seem to blur the boundaries between the *Nightmare on Elm Street* film series and reality itself.

The narrative begins with Langenkamp being approached to star in a new *Nightmare* film, ten years after the franchise's apparent conclusion with *Freddy's Dead: The Final Nightmare* (1991). As she deliberates this decision, strange occurrences begin affecting her family life: her husband Chase, a special effects artist, dies in a car accident under mysterious circumstances involving apparent claw marks, while her young son Dylan begins exhibiting disturbing behavioural changes and an obsession with the character of Freddy Krueger.

Through conversations with Craven himself, who appears as a character within his own film, Langenkamp learns that the *Nightmare* films served to contain an ancient demonic entity that feeds on "the murder of innocents." This entity, having been temporarily bound by the film series, has now escaped due to the franchise's conclusion and seeks to cross over from the realm of fiction into reality, using Freddy Krueger as its preferred manifestation. The film suggests that this entity can only be re-contained through the creation of a new story powerful enough to capture its essence, which becomes the very film we are watching.

The climactic sequences see Langenkamp forced to embody Nancy Thompson once more, this time in the "real" world, as she battles the entity through a landscape that merges her actual Los Angeles home with the surreal geography of Freddy's dream realm. The film concludes with an earthquake that literalizes the collapse between fictional and actual reality, as Langenkamp must use her knowledge of horror film conventions to defeat an enemy that has transcended its original fictional boundaries.

We argue that this film functions as a theory-constructing thought experiment, as it aims to build a new theoretical perspective, specifically what we can call Craven's "Media Archetype Theory". The film makes claims about how fictional characters can become real through collective belief, how censorship backfires, how commercialisation corrupts stories, and how filmmakers are responsible for what they create. However, rather than presenting this theory through traditional philosophical explicit argumentation, the film builds it systematically through its metatextual structure, demonstrating each component through cinematic techniques. The film represents "every single sentiment that Craven ever expressed about why we tell horror stories, all of his scholarly knowledge in myth and storytelling, that is all here, distilled into a two-hour runtime" (Brehmer 2019).

We should note that it can be challenging to defend that a film can perform a theory-construction thought experiment. Contrary to simpler thought experiments that establish single points, such as counterexamples or possibility-establishing ones, theory construction involves a systematic development of different interconnected claims. This creates a kind of explicitness escalation problem: the more philosophical complexity we claim for a film, the more it appears to require explicit articulation of premises. While a film might suggest individual philosophical insights, theory construction requires that the audience understand not just isolated points, but also their theoretical relationships. This can lead to a version of the dilemma of the paraphrase. If the film requires explicit articulation to demonstrate these logical connections, then the philosophical work is being done by the paraphrase rather than the cinematic experience itself. Conversely, if the film works purely through cinematic means without such articulation, it may fail to establish the systematic interconnections that characterise theory construction thought experiments. The challenge, then, is to demonstrate how a film can achieve the coherence of a theory-constructing thought experiment while remaining distinctively cinematic.

A solution may lie in an analogous case found in traditional philosophy: Plato's *Republic*. The *Republic* may be viewed as a theory-construction thought experiment that incorporates other

nested, distinct types of thought experiments, Socratic dialogue, extended metaphors, allegorical narratives, and mythological accounts. Despite this diversity, readers gain a clear understanding of Plato's theory of justice, as harmony between parts in both the soul and the state, without the need for an explicit argument stating this as a formal thesis. The coherence of the theory stems from the integration of multiple approaches and points, allowing readers to construct a systematic understanding themselves. I may argue that a film can achieve a similar effect by employing multiple cinematic techniques and incorporating nested thought experiments. The film would enable audiences to understand the systematic construction of theory by integrating diverse cinematic elements. I will try to show how this can be the case for this film.

6.5.1 Evil archetypes and the making of meaning

The film progresses through a series of steps, integrating interconnected philosophical domains that build each component of the theory-construction thought experiment.

The basic premise that the film establishes is that there is a relationship between archetypal patterns and collective cultural engagement. On the one hand, it posits that archetypal forces exist as pre-existing psychological and cultural potentials. This might be understood through Carl Jung's concept of the collective unconscious, which Craven himself references when noting that "Jung would have been more approving of me than Sigmund Freud" (Craven cited in Shapiro 1994). These archetypal patterns constitute what Jung describes as an "archaic or – I would say – primordial types, that is, with universal images that have existed since the remotest times" that form part of the collective unconscious, which "has contents and modes of behaviour that are more or less the same everywhere and in all individuals" and "constitutes a common psychic substrate of a suprapersonal nature which is present in every one of us" (Jung 1969). They represent persistent human psychological structures that manifest across various cultural contexts and historical periods. Through Craven's character, the film establishes this premise: certain fictional narratives serve as containers for what he calls (in the film) "an ancient demonic entity that feeds on 'the murder of innocents'" that "has been given many different names through the years by many different storytellers." Craven himself explicitly connected his work to these psychological foundations, stating: "I never did anything to rewrite horror; I just tried to do something interesting... I had no hesitation about writing something parallel to what I was studying at the time – Eastern religions and meditation, Sufism" (Craven 2015).

On the other hand, certain fictional narratives serve as containers for archetypal forces that have "been given many different names through the years by many different storytellers" (Craven in *New Nightmare*). The story suggests that when communities engage collectively with particular narrative representations of archetypal patterns, these representations can gain an autonomous cultural existence that transcends their original fictional boundaries. Thus, there is a transformation of archetypal potential into active cultural forces that, in their turn, start influencing reality beyond their narrative containers, requiring new stories to re-contain them.

New Nightmare constructs an archetypal media theory through the following propositions:

1) Fictional narratives contain psychological forces that exist across cultures;
2) When audiences engage collectively with narratives, these forces can become real and break free from their original boundaries;
3) When narratives become corrupted through commercialisation or suppressed through censorship, contained archetypal forces "break free" and require new stories for re-containment;
4) Cinema functions as both a medium that can release archetypal forces and the necessary technology for their cultural re-containment;
5) Therefore, filmmakers bear cultural responsibility for maintaining authentic archetypal engagement rather than corrupted commercial exploitation.

The opening sequence exemplifies this process through the mechanical construction of Freddy's glove. We begin with the mechanical construction of Freddy's glove as a film prop. But after some autonomous movement, the glove begins to operate independently, eventually attacking the crew. After this, Heather wakes up abruptly (because of an earthquake that we later find out was caused by the emergence of a "real" Freddy) from what we now know is a nightmare.



Figure 39: Still images - Opening scene with Heather's dream sequence.

This demonstrates the transformation process that the theory describes: sustained collective engagement (as seen in the decades of Nightmare films) has transformed what was originally just a prop into something with quasi-autonomous existence. While it may seem at first to be just a simple nightmare, Freddy and his horror are emerging in reality.

New Nightmare employs earthquakes as a literal cinematic manifestation of archetypal forces breaking through fictional boundaries into reality. Craven times these earthquakes with Heather's nightmares about Freddy: when she dreams about him, the earth shakes. This shows that her engagement with Freddy is causing real-world effects. The film presents earthquakes through sudden camera shaking and environmental disruption, allowing viewers to experience these events simultaneously with the characters, making the collapse between fictional and real feel immediate rather than abstract. We do not just think about fictional characters becoming real: we see the ground shake when it happens.

This film, thus, shows how archetypal forces persist across different cultural periods while manifesting through historically specific forms: "Evil takes many forms in many times, because the stories we tell to scare ourselves are always the same, at their core. It was the witch in 'Hansel and Gretel.' It was Count Dracula, or Count Orlok from Nosferatu at the turn of the century. It was the cave bear that early man had nightmares about. And in the 1990s, it is Freddy Krueger" (Brehmer 2019).

New Nightmare's climactic resolution shows this component of the theory, too. When the archetypal evil breaks free from its narrative containment (due to the *Nightmare* series ending), Heather Langenkamp must consciously choose to embody Nancy Thompson, not as an actress playing a role, but as herself channelling the archetypal "final girl" pattern that Nancy represents. This is important because it shows the theory working in both directions: the same collective cultural engagement that allowed the evil to achieve autonomous existence also provides the mechanism for re-containment. Heather cannot defeat this "real" evil on her own; she needs to access the archetypal power that Nancy embodies through decades of collective audience engagement. However, she also cannot lose herself in the role; she must consciously choose to channel Nancy while remaining her true self, Heather. This shows that fictional characters can become real through collective belief. Nancy is not just a movie character anymore, as she has become a real cultural force that Heather can tap into.

Heading towards the final scene, Heather begins developing Nancy's physical characteristics, most notably a white streak appears in her hair, identical to Nancy's. This is a manifestation of how sustained engagement with *Nightmare* narratives blurred the boundaries between actor and character.



Figure 40: Still images - “Nancy” in *Nightmare* (1984) and “Heather” in *New Nightmare* (1994)

When Heather enters Freddy’s lair to rescue Dylan, the film demonstrates the archetypal theory’s claims about cross-cultural manifestations. She encounters traditional symbols of evil, like snakes, that connect Freddy’s contemporary horror to ancient archetypal representations of malevolence.

Then Heather discovers Craven’s script within Freddy’s domain. The script’s presence in the evil’s space literalises the claims about the interpenetration of artistic creation and autonomous reality: the boundary between Craven’s creative work and the “real” evil it contains has completely dissolved.

Heather’s discovery of Craven’s script within Freddy’s domain represents one of the film’s most radical claims about the relationship between artistic creation and autonomous reality. The script’s presence in the evil’s space demonstrates that the boundary between creative work and the forces it seeks to contain has completely dissolved. This is not merely symbolic but ontological: Craven’s artistic creation has become literally interpenetrated with the archetypal evil it was meant to contain. The script functions as evidence that the creative act itself participates in the reality it describes rather than standing outside it as mere representation. When fictional narratives become real through collective belief, the creative process gets tangled up with the forces it creates. Artistic creation and evil are not opposites; they are working together in the same process. This challenges assumptions about artistic distance and creative control, revealing that when archetypal forces escape their containers, they take the creative process itself as part of their domain

The sequence explicitly connects to the Hansel and Gretel archetype through multiple elements: Heather and Dylan use pills as breadcrumbs to navigate the evil space, and Freddy literally tries to devour Dylan.

Heather's movement through the floor, walking down steps that mirror the iconic steps sequence from the original *A Nightmare on Elm Street*, shows archetypal patterns repeating across the fiction/reality boundary.

The metatextual structure of *New Nightmare* creates the conditions under which viewers understand that they actively participate in the making of meaning, rather than simply being passive recipients of meaning. The film accomplishes this through what Craven described as an investigation into "the various states they're in while watching a movie, states they might not ordinarily recognize" and his argument that "watching television or movies is like being unconscious" (Craven 1990). In this regard, the film is also designed to awaken viewers to consciousness of their interpretative processes. Through this "awakening", viewers do not simply learn about Craven's archetypal media theory: they participate in its construction.



Figure 41: Still images - The stairs in *Nightmare* (1984) and in *New Nightmare* (1994).

Finally, the defeat of Freddy through fire, burning him in a furnace, connects to traditional methods for destroying witches and evil entities across cultures. This is not just a convenient plot resolution but a demonstration that the same archetypal solutions that have contained evil throughout human history remain effective even when those forces achieve autonomous existence through modern media engagement.

The entire sequence functions as proof that archetypal patterns operate through but transcend specific cultural manifestations, that fictional narratives can achieve autonomous reality through collective engagement, and that the same mechanisms that allow archetypal forces to escape their containers also provide the means for re-containment.

6.5.2 Censorship and authorial responsibility

The film also develops a critique of censorship through its demonstration that repressing cultural narratives causes archetypal forces to "break free" rather than containing them.

The character of Dr. Heffner embodies these censorship practices. The character is named after MPAA (Motion Picture Association) chairman Richard Heffner, whom Craven viewed as a tyrannical censor, calling him "the kind of man who will not allow a certain kind of view to be expressed, period" (Craven, cited in Mitchell 1990).

Dr. Heffner immediately connects Dylan's disturbing behaviour to Heather's involvement in the *Nightmare* series: "You haven't showed him any of the films you've made, have you?" The film positions such "protection" as more threatening than the material it seeks to suppress, demonstrating how censorship functions as a form of violence against necessary cultural processes.

When Dr. Heffner appears in Dylan's nightmare wearing Freddy's bladed glove while declaring her intention to "cut the evil out" of Dylan, the film visually demonstrates its claim that "It is the telling of these stories that processes that fear instead of repressing it. When repressed it grows restless and breaks free" (Brehmer 2019). This sequence constructs the anti-censorship component of the theory.



Figure 42: Still image - Dr. Heffner, “Let’s cut this evil out of him!”

Craven revealed this premise in an interview: “We were in the middle of dealing with a lot of strict censorship. We asked ourselves what these films give or take away from culture [...] It’s like boot camp for the psyche. This is the way humans deal with the horrors of existence. If you forbid this kind of art, the actual, real horror is unleashed in a sense” (Craven 2015).

As an alternative solution to outright censorship, the film proposes that creators should bear some responsibility for the cultural impact of their work, while also acknowledging their inability to fully control how their creations are interpreted and utilised. Creators cannot disclaim responsibility for the cultural effects of their work while maintaining artistic integrity; yet, they also cannot predict or prevent all possible interpretations and applications of their work.

When we see Craven’s script on his computer showing the exact conversation he is having with Heather, it is unsettling. The film shows that creators cannot fully control what they release, but they are still responsible for creating the conditions that let it happen.

Craven’s concerns about authorial responsibility extend beyond censorship to include his critique of how commercial forces appropriated and corrupted his original creation. In interviews, he expressed frustration with how the *Nightmare* sequels abandoned the philosophical foundations of his work: “I would have hoped that the *Elm Street* films would have been treated with absolute respect along the way [...]. I would have liked to have seen somebody sit down each time they set out to make one of the sequels and really get into the philosophy and the heart behind it” (Craven cited in Shapiro 1994).

This critique highlights an important aspect of the authorial responsibility paradox: creators lose control not only to censorship but also to commercial exploitation that undermines the cultural

function of their work. Craven argued, "My first film was about some very serious and important subjects. I felt that with 2, they immediately threw all of the important issues out the window and made it a series of strange, freaky events and the same old raunchy teenagers" (Ibid.). The sequels reduced complex archetypal material to formulaic entertainment, exactly what *New Nightmare* theorises happens when narratives "get too familiar to people, or somebody waters it down to make it an easier sell" (Craven in *New Nightmare*). The film demonstrates this corruption through its portrayal of how Freddy devolved from "pretty much the embodiment of evil" in the original to "a veritable comedian in his later movies" where "you almost want to hang out with the guy" (Nunez 2022). This transformation represents more than poor filmmaking; it constitutes a failure of cultural responsibility that allows archetypal forces to escape their narrative containers.

New Nightmare positions commercial appropriation as a form of cultural violence parallel to censorship, as both prevent narratives from fulfilling their essential archetypal work. Where censorship suppresses necessary cultural engagement, commercialisation corrupts it by prioritising market appeal over authentic archetypal content. Both create conditions where archetypal forces can no longer be properly contained within their narrative frameworks, forcing creators like Craven to reclaim and reconstitute their original vision to restore the cultural work that commercialised versions have abandoned. This connects the film's anti-censorship stance to its critique of market-driven content creation, showing how prohibition and commercial corruption threaten the cultural function of archetypal narratives.

6.5.3 Philosophical synthesis and legacy

Although it was not his final film, and it precedes *Scream*, *Wes Craven's New Nightmare* represents the culmination of philosophical concerns that developed throughout his career. It achieves something more ambitious than mere synthesis: it creates a unified theory that resolves tensions and contradictions that had remained unresolved in his earlier films. Where previous works explored isolated philosophical questions, *New Nightmare* integrates these explorations into what one critic recognises as "his master's thesis on a boogeyman he created, and loved, but could never escape, as well as the concept of boogeymen, in general" (Brehmer 2019). This achievement represents not only the summation of earlier insights but also their systematic integration into a theory that validates the philosophical potential that had been developing throughout Craven's filmography.

New Nightmare brings together philosophical themes from across Craven's career. *The Last House on the Left* (1972) explored questions about violence and moral boundaries; *A Nightmare on Elm Street* (1984) investigated the relationship between consciousness and reality; *Scream* (1996) examined media effects and cultural reflexivity, but appeared after *New Nightmare* and, thus, benefits from the theoretical groundwork the earlier film established. *New Nightmare* accounted for how violent representation functions across different cultural contexts while maintaining ethical responsibility.

The film accomplishes something important here. It resolves the tensions between popular entertainment and serious philosophy - a persistent problem that we can trace throughout Craven's career. As the film shows, "Every kid knows who Freddy is. He's like Santa Claus, or King Kong" (Langenkamp in *New Nightmare*). This cultural penetration gives popular narratives a philosophical reach that academic discourse cannot match. The film's archetypal media theory suggests that cultural narratives must achieve popular resonance to perform their essential psychological and philosophical work.

This validates Craven's belief that horror can do serious cultural work while remaining popular. Some philosophical ideas need cinema to work as they cannot be done as well through text alone.

The film does not just show that film can philosophise: it shows that some philosophical ideas need cinema. We can see that the film's archetypal media theory requires cinematic development rather than purely textual means – although we could achieve it through those means, it would be impoverished. This makes sense when we consider what the theory examines: how the cinematic medium functions as a kind of philosophical technology.

When we notice the pills work like breadcrumbs in *Hansel and Gretel*, or connect Dr. Heffner to censorship, we are actively building the film's meaning. This proves the film's point about how cultural meaning works. This collaborative process corroborates what the film claims about cultural meaning. Rather than emerging through passive consumption, we see that cultural meaning emerges through our active engagement. The film shows how horror stories help us think about and deal with evil through popular culture that everyone can access.

6.6 Summary and final considerations

This chapter aimed to demonstrate that Wes Craven's films function as philosophical thought experiments, employing distinctively cinematic techniques to investigate fundamental questions about violence, reality, representation, and human nature.

Having analysed four of Craven's films as philosophical thought experiments, I can now evaluate how successfully these cases fulfil the previously established criteria and methodological approach.

When we consider criterion *a* (interpretability within philosophical conversations), we can see that each Craven film engages with established philosophical discourse. At the same time, these films contribute fresh perspectives through their cinematic presentation. The *Last House on the Left* directly engages with moral philosophy's longstanding debates about revenge, retribution, and justice. We can trace how the film engages with established ethical discourse. It connects to Aristotelian theories of justified anger and contemporary discussions about restorative versus retributive justice. However, as we have seen, the film also challenges something important: our assumptions about moral boundaries when circumstances become extreme.

A Nightmare on Elm Street systematically explores fundamental epistemological questions about the reliability of perception and the nature of consciousness that trace back to Cartesian scepticism and Platonic metaphysics. *Scream* engages with multiple philosophical traditions simultaneously: media theories, moral psychology (challenging assumptions about media effects on behaviour), and metaphysics (exploring questions about simulation and representation). The film's metacinematic approach connects to established philosophical discussions on the nature of cinema in general, and horror cinema specifically. *New Nightmare* constructs what we termed an "archetypal media theory" that relies on Jungian psychology, media studies, and aesthetic theory to explore questions about authorial responsibility, cultural meaning-making, and the ontological status of fictional entities. Even its most innovative claims, about narratives containing archetypal forces, build on established theoretical foundations.

If we examine criterion *b* (following the structure of philosophical thought experiments), we can see that my case studies demonstrate diverse approaches. These approaches mirror the methodological structure we find in traditional thought experiments.

The Last House on the Left functions as a counterexample by presenting what initially appears to be a maximally justified case for revenge (parents responding to their daughter's brutal murder) and then systematically demonstrating how even this "ideal" scenario leads to moral corruption. The film's structure mirrors classical counterexample methodology: establish

conditions that should validate the target theory, demonstrate that these conditions produce contrary results, and therefore challenge the theory's general validity. *A Nightmare on Elm Street* operates as a possibility-establishing experiment that systematically explores "What if the boundaries between dreaming and waking consciousness became unstable?" The film does not simply assert this possibility but develops its implications through Nancy's escalating confusion about which experiences count as "real". Each scene builds toward the central insight that our ordinary criteria for distinguishing reality from illusion may be unreliable. *Scream* functions as multiple thought experiments simultaneously: it establishes the possibility of hyperreality (where media saturation collapses fiction-reality distinctions), provides counterexamples to simplistic media effects theories (by showing characters who consume violent media but respond appropriately to real violence), and tests hypotheses about genre conventions and audience expectations. *New Nightmare* represents the most complex case, functioning as a theory-constructing thought experiment that integrates claims about fictional ontology, cultural meaning-making, and authorial responsibility into a unified "archetypal media theory".

Concerning criterion *c* (the philosophical understanding emerges through the cinematic experience itself), my analysis demonstrates that philosophical insights emerge through the cinematic experience itself rather than requiring post-viewing extraction. The visceral impact of *The Last House on the Left* forces viewers to feel the moral equivalence between victims and victimisers rather than contemplating it abstractly. The film's philosophical work occurs through the viewing experience's emotional and cognitive disruption. *A Nightmare on Elm Street*'s "rubber reality" effects create epistemological uncertainty in viewers that mirrors Nancy's experience. When the film presents Krueger's hat materialising in the physical world, viewers experience rather than merely consider the breakdown of categorical boundaries between dreams and reality. *Scream*'s layered visual composition and self-reflexive editing create a vertigo that makes viewers experience the impossibility of distinguishing between representational levels. The philosophical insight about hyperreality emerges through this direct experiential disorientation. *New Nightmare*'s metatextual structure guides viewers through the systematic construction of its archetypal media theory by making them active participants in the collapse between fictional and actual reality. The philosophical theory emerges through collaborative engagement with the film's cinematic presentation.

Regarding my methodological proposal, my analysis applied mental modelling and constructivist approaches throughout. *Last House on the Left* activates "chunked material" from

viewers' experiences of family relationships and moral decision-making, reorganising these elements to reveal hidden connections between revenge and corruption through Gendler's "quasi-observational" process; *Nightmare*'s manipulation of temporal and spatial continuity facilitates mental modelling that integrates explicit content with tacit knowledge about consciousness and reality, activating cognitive resources that textual descriptions cannot access; *Scream*'s layered composition activates viewers' stored knowledge about genre conventions and reality-fiction boundaries, reorganising cognitive schemas to reveal how media saturation collapses categorical distinctions; *New Nightmare*'s metatextual structure engages viewers' tacit knowledge about storytelling and authorial responsibility, systematically reorganising these elements into a unified archetypal media theory.

Our analyses also identified distinctively cinematic techniques and how they enable the philosophical work being done. *Last House*'s parallel framings, elimination of anticipated heroic cinematography, clinical presentation of violence; *Nightmare*'s editing patterns and visual effects that create the epistemological conditions the film investigates, creating uncertainty about reality; *Scream*'s complex framing, rapid cross-cutting between levels and the use of film within film generate a phenomenological experience of hyperreality; *New Nightmare*'s metatextual techniques, particularly the use of "real" actors playing themselves, the earthquake sequences that literalise the collapse between fictional and actual realities, and the visual parallels with *Hansel and Gretel* with the film's narrative constitute its archetypal media theory. The film's mise-en-scène (showing "real" Los Angeles alongside surreal dream geography) creates ontological uncertainty about fictional entities.

The interpretative approach I took balanced Craven's documented intentions with cinematic evidence and viewer engagement. I supported Craven's explicit statements in interviews with specific moments in the films, and vice versa. My collaborative model recognised both filmmaker architecture (Craven's construction of thought experimental scenarios) and viewer interpretation (audience engagement with the philosophical implications). This approach avoided both the imposition objection (by grounding interpretation in textual evidence and authorial context) and reductive intentionalism (by acknowledging the role of viewer collaboration).

We applied traditional thought experiment categories, and it revealed that certain types of philosophical thought experiments are particularly well-suited to the cinematic medium. Counterexample thought experiments, as demonstrated by *The Last House on the Left*, leverage cinema's visceral impact to challenge moral intuitions. The film's portrayal of violence forces

viewers to confront the inadequacy of simple revenge positions. Possibility-establishing thought experiments, exemplified by *A Nightmare on Elm Street*, exploit cinema's unique capacity to make impossible scenarios feel plausible through visual presentation. The film's "rubber reality" depends essentially on cinema's manipulation of temporal and spatial continuity in ways that purely textual description cannot replicate. When Nancy pulls Krueger's hat from her dream into physical reality, viewers experience the breakdown of categorical boundaries rather than merely contemplating it. Theory-constructing thought experiments, while more complex to defend, prove achievable through cinematic means, as *New Nightmare* demonstrates. Like Plato's *Republic*, which integrates multiple approaches without the need for explicit thesis statements, Craven's film constructs its archetypal media theory through the systematic integration of diverse cinematic elements. The film enables audiences to understand theoretical relationships through interpretive collaboration rather than explicit exposition of verbal arguments.

Other types of philosophical thought experiments remain largely absent from my analysis, either because they prove less amenable to narrative cinema or because they do not appear prominently in Craven's work. Demonstrating impossibility seems to be the most challenging type for cinema to achieve. Such thought experiments typically work by revealing logical contradictions or conceptual incoherence through abstract reasoning. Cinema's concrete nature makes it better suited to showing practical difficulties or implausible scenarios rather than genuine logical impossibility. Confirming theories types represent perhaps the most accessible type of thought experiments for films generally, and Craven's work provides numerous examples. His movies consistently support various theories, including psychoanalytic theories about repression and the return of the repressed (Nightmare), anthropological theories about the ritual processing of cultural trauma (*Last House*), media effects theories (*Scream*), and Jungian archetypal psychology (*New Nightmare*). Indeed, *New Nightmare* serves as a particularly complex case, simultaneously confirming archetypal theory while providing counterexamples to simplistic claims about media effects and establishing new possibilities regarding the relationships between fiction and reality. Establishing necessary connections is theoretically possible for cinema, but remains largely absent from my analysis. Such thought experiments demonstrate that if *x* obtains, then *x* must necessarily follow through a logical connection. While films excel at showing patterns and causal relationships, they lack the abstract logical work required to establish strict necessity rather than strong probability. However, my analysis does not exclude the possibility that cinema could successfully perform these unexplored types of thought experiments. Future research on other films and

filmmakers, or these films from a different perspective, might reveal effective cinematic strategies for the remaining types of thought experimentation.

My focus on a single filmmaker's coherent body of work, rather than surveying multiple directors to locate examples of each thought experiment type, reflects a deliberate methodological choice. Craven's career-spanning philosophical project offers good insight into how cinema can systematically investigate philosophical questions.

Importantly, my argument for films as philosophy does not depend on claims about artistic excellence. The philosophical value of Craven's work stems from its effective deployment of cinematic techniques to enable philosophical investigation, not from its status as "great cinema." The artificial separation between good cinema and serious thought becomes evident in recent criticisms of films like *Prometheus* (2012) and *Alien: Covenant* (2017), which some fans dismissed as "too philosophical" and therefore poor cinema. Such reactions reveal the false dichotomy that my analysis challenges. However, films must deploy cinematic techniques competently enough to enable the cognitive work that philosophy requires.

Recognising Craven's films as philosophical investigations reveals a substantial body of work that has been overlooked. His exploration of revenge and violence offers alternatives to retributivist frameworks, suggesting that moral theory must account for the psychological and social contexts that shape human responses to trauma. His investigation of consciousness and reality anticipated contemporary debates about simulation theory and virtual reality, while providing grounded analyses of experience that abstract philosophical discussions lack.

Perhaps most significantly, Craven's archetypal media theory provides essential insights into how contemporary culture processes collective anxieties through popular narratives. His analysis of evil archetypes reveals fundamental aspects of human nature and cultural psychology. In Craven's films, horror serves as a form of cultural therapy, allowing societies to confront and contain fears. Thus, my main argument here is that Craven philosophised not through textual explicit means, but through cinematic thought experimentation.

Table 2: Case Studies Analysis - Films as Philosophical Thought Experiments

Film	The Last House on the Left (1972)	A Nightmare on Elm Street (1984)	Scream (1996)	New Nightmare (1994)
Type of Thought Experiment	Counterexample	Possibility-Establishing	Possibility-establishing, Counterexample	Theory-Constructing
Philosophical Question	Are there circumstances that morally justify personal revenge?	What if boundaries between dreaming and waking consciousness became unstable?	Does media cause real violence? What happens when fiction/reality boundaries collapse?	How do fictional narratives gain cultural power? What is the relationship between authors and their creations?
Cinematic Techniques	Visceral violence presentation, parallel editing	Dream sequences (editing), materialisation of objects (props)	Metatextual elements, dialogue, editing Genre rule articulation, mise-èn scene, voice modulation	Metatextual elements, dialogue, dream sequences, genre rule articulation, mise-en-scéne, props, visual metaphor and parallel
Philosophical Contribution	Even "maximally justified" revenge leads to moral corruption	Our ordinary criteria for distinguishing reality from illusion may be fundamentally unreliable	Media is not responsible for violence. Hyperreality emerges through media saturation.	Narratives contain autonomous forces that can influence reality through cultural circulation

7. Metaphilosophical considerations

Throughout this thesis, I have examined various objections to the film as philosophy hypothesis. I explored the banality, the explicitness and the generality objections, which posit that film's cognitive contributions are trivial, demand that philosophy produce explicit argumentation and argue that philosophy must establish universal truths. I investigated how the appeal to thought experiments might address these objections, examining how films could function as philosophical thought experiments through cinematic means.

My journey continued through an analysis of the rationality objection and Livingston's bold thesis, which challenged whether films could make genuinely distinctive philosophical contributions. I addressed concerns about authorial intention and the imposition objection, developing a moderate intentionalist framework for interpreting cinematic philosophy. Ultimately, through case studies, I demonstrated how specific films can serve as distinct types of philosophical thought experiments, utilising cinematic techniques to explore philosophical questions in ways that complement traditional philosophical methods.

As my investigation has progressed, I have noticed a pattern: each pro-film as philosophy argument and objection rests on specific assumptions about what philosophy *is* and how it *should be* conducted. Bruce Russell's insistence that philosophy requires explicit argumentation reflects a particular metaphilosophical commitment to formal reasoning as the essence of philosophy. Murray Smith's emphasis on explicitness, empirical support and reflective maturity embodies a normative view of philosophical practice that privileges specific methods over others. Even more pragmatic concerns, like the rationality objection, depend on prior judgements about what counts as philosophically valuable.

This suggests that the film as philosophy debate cannot be resolved by examining cinema's capacities alone. It also requires us to confront questions about philosophy itself: What are philosophy's essential features? Which methods should we count as philosophical? How broadly or narrowly should we draw the boundaries of philosophy?

The methodological implications are also fundamental: the question becomes whether we should start with what we think philosophy is and see if films can correspond to that, or we look at films and see if philosophy corresponds to that. This choice between starting with philosophical strict philosophical preconceptions or remaining open to what cinema might teach us about

philosophical thinking reflects broader questions about how philosophy relates to other disciplines and cultural practices.

Thus, the film as philosophy debate illuminates tensions within philosophy between descriptive and prescriptive approaches to understanding it. A descriptive metaphilosophy might simply catalogue what philosophers have done historically, potentially making the boundaries of philosophy so broad as to include almost any thinking. A prescriptive approach might establish rigid criteria for philosophy, possibly excluding valuable and even canonical ways of philosophising. The challenge may lie in finding a middle ground that maintains philosophical standards while not excluding traditional work and remaining open to innovation.

This chapter argues for a moderate metaphilosophical position that can resolve the film as philosophy debate without falling into either excessively restrictive or inclusive ideas of philosophy. Based on the critiques I developed of narrow metaphilosophical approaches and the positive case I built for cinematic thought experiments, I will defend a conception of philosophy that is open to new modes of expression while maintaining standards for what counts as valuable philosophical contribution.

The argument will follow three main steps:

1) How restrictive metaphilosophies create artificial barriers to film's philosophical potential, often by defending essentialist accounts that exclude too much recognised philosophical work;
2) How overly inclusive and alternative approaches risk making philosophy indistinguishable from any form of thinking;
3) Defence of a moderate account of philosophy that emphasises case-by-case analyses rather than general statements about the relationship between film and philosophy.

7.1 The problems of essentialist metaphilosophies

According to Søren Overgaard, Paul Gilbert, and Stephen Burwood, essentialist accounts attempt "to lay down a definition, an eternal fence, so what lies within is philosophy, and what lies without is not" (Overgaard et al. 2013, 20). Such approaches typically define philosophy either topically (in terms of the "big questions" it addresses) or methodologically (in terms of specific procedures like conceptual analysis). However, a metaphilosophy that focuses only on one and not the other may

be problematic: “an account of philosophy as conceptual analysis excludes too much, and an account of philosophy as thinking about the big questions includes too much” (Ibid., 20-22).

Murray Smith argues that supporters of film as philosophy usually adopt an “expansive” or a “reductive” strategy. The first “begins with a looser, more inclusive conception of philosophy,” and the latter “around a specific, narrow conception of philosophical activity” (Smith 2006, 34). What Smith seems to say is that when defending or objecting to film as philosophy we need to defend specific conceptions of philosophy. This may be because the acceptance of a looser conception of philosophy facilitates the idea that every film can philosophise, and that that might not be a bullet we would be willing to bite. In such a case, how would we decide if a film is philosophy or not, or if it is bad or good philosophy? That is only possible to decide when we defend “prescriptive” or normative metaphilosophy, i.e., if we state how philosophy should be done.

Philosophy, like other disciplines, has evolved throughout its history. The methods, questions, and standards that define contemporary philosophy differ from those employed by pre-Socratics, ancient Stoics, or medieval scholastics. Attempting to identify strict essential features risks either excluding recognised philosophical traditions or defining philosophy so broadly as to lose any boundaries.

Furthermore, Overgaard et al. suggest how to confront and evaluate each metaphilosophical position:

1) A metaphilosophy should not be too revisionist, meaning it should not “imply that most philosophers (past and/or present) have been doing philosophy in completely the wrong way”;
2) it should “explain the relative ‘lack of progress’ characteristic of philosophy relative to the natural sciences” (Ibid., 25).

To illustrate the revisionist problem, consider Overgaard and colleagues’ analogy with football: “consider someone who suggested that football would be better played without goalkeepers. One’s reply would be that this might be an interesting game (...) but it would not be football” (Ibid., 25).

For example, Russell and Smith are essentially arguing that philosophy would be better if it excluded ambiguous, non-empirical, or non-argumentative approaches. However, this suggests creating a new intellectual discipline rather than accurately characterising the historical and contemporary practice of philosophy.

Both Russell's and Smith's approaches exemplify broader problems with essentialist metaphilosophy. Restrictive metaphilosophies might be motivated by professional protectionism. As Wartenberg says, "philosophers are, after all, jealous of their professional identities" (Wartenberg 2007, 132).

Regarding the lack of progress issue, the idea that philosophy does lack progress is debatable. Eric Dietrich proposes a thought experiment in order to understand why, according to him, philosophy does not progress.

> Imagine that Aristotle [...] pops forward to today, on a well-known campus somewhere in some English-speaking country [...]. Curious about the state of knowledge, he finds a physics lecture and sits in. What he hears shocks him. A feather and an iron ball fall at the same rate in vacuum [...]. Sadly, he trundles off to a philosophy class – a metaphysics class, as it turns out. [...] He knows exactly what the professor is talking about. [...]. Then Aristotle goes on an ethics class, where he learns of the current importance of what is apparently "virtue ethics". He recognises it immediately [...] (Dietrich 2011, 334).

This seems to be an indication that, contrary to science where, for example "Darwin's conclusions are regarded as *true*," on the other hand in philosophy we still do not know whether a philosopher's theory is right or not, for example "Mill 's conclusions are not regarded as true, but rather, interesting and important" (Dietrich 2011, 335). Now, as Dietrich argues, this is not to say that philosophy does not transform somehow, especially in its different methods or technical language; however, it seems undeniable that philosophy at least does not progress into an agreement about propositions as science does. Thus, it should be said that the mentioned lack of progress to be explained is relative, compared to the sciences.

7.1.1 Restrictive metaphilosophies

As we have seen, the most prominent objections to film as philosophy rest on restrictive conceptions of what philosophy is and how it should be conducted. These approaches typically implicitly adopt essentialist strategies that define philosophy by identifying its supposedly necessary and sufficient conditions. By examining the metaphilosophical foundations underlying the explicitness and generality objections, we can see how these restrictive approaches create problems not only for film's philosophical aspirations but for philosophy itself.

Bruce Russell's objection to film as philosophy exemplifies the problems with essentialist metaphilosophy. As we have seen, his argument follows a logical structure that reveals his underlying assumptions about philosophy's nature:

1) Philosophy involves justifying beliefs about general philosophical claims. Because answers to philosophical questions are not obvious, explicit argumentation is necessary for this justification.
2) Films do not provide explicit argumentation.
3) Thus, films cannot justify beliefs about general philosophical claims.
4) Therefore, films cannot philosophise (Russell 2008, 20).

This argument embodies a kind of methodological essentialism, the view that philosophy is essentially defined by its use of explicit argumentation.

Russell's approach faces important problems. He assumes without argument that explicit argumentation is necessary for philosophical work. So, immediately, his claim could be accused of being ambiguous or lacking explicit argumentation for it. In the same vein, Wartenberg argues that Russell does not clarify what films specifically have, or lack, that makes them unable to produce arguments, presenting several possibilities: is it their visual nature, or the narrative structure? We can infer that the implied assumption is that only linguistic expression, verbal or written, can produce this kind of argumentation. If films cannot provide, then Wartenberg asks, "Is there ambiguity in Russell's grounds for this assertion? Does this mean he is not making a philosophical argument?" (Wartenberg 2008, 32).

He argues that the connection between films and thought experiments is irrelevant because "thought experiments can be a stimulus to argumentation, can motivate the listener, reader, or viewer to produce an argument, but they are not arguments themselves" (Russell 2008). Furthermore, he concedes that thought experiments "can provide non-inferential justification of claims," which can, then, "serve as data to defeat proposed analyses of some concept", but that data needs explicit argumentation (Ibid.). Then he concedes that "film can provide the same sort of data that philosophical thought experiments provide, but because it cannot provide the explicit argumentation, except by simply 'downloading an argument onto the soundtrack,' it cannot yield justified beliefs about general conceptual propositions" (Ibid.). Thus, here he only refers to thought experiments that function as counterexamples, not any other types, and even in this case, he argues that even if films can be equated to this type of thought experiment, it is not enough to be considered philosophy as it does not include the explicit argumentation philosophy requires.

Russell's (implied) essentialist approach of philosophy as requiring explicit argumentation has the consequence of excluding too much recognised philosophical work. For instance, Wartenberg refers to the historical dispute over how to characterise the precise conclusion of the Transcendental Deduction of Categories, in Kant, concluding that "just because an argument is philosophical, it need not be unambiguous" (Wartenberg 2007, 20). Carroll also appeals to the existence of dialogues or texts in which knowledge is presupposed to the readers' background, so that they understand the argument and the thesis without needing all the premises explained (Carroll 2008, 20). He recalls that there are philosophers who have expressed themselves through literary devices, such as rhetorical questions, aphorisms, and fictional narratives (particularly thought experiments), arguing that Russell's type of philosophical presupposition "will not be accepted by anyone who regards Nietzsche's aphorisms or genealogies or Wittgenstein's puzzles as philosophy" (Carroll 2017, 277).

Russell might respond that these counterexamples assume a deflationary or institutional definition of philosophy that simply accepts whatever has been historically recognised as philosophy, one that argues that philosophy "is whatever the people who are employed *as* philosophers at universities and other institutions do, or whatever material librarians catalog as such – end of story" (Overgaard et al. 2013, 20).

However, this misses the fact that Russell provides no independent justification for why explicit argumentation should be at least one of philosophy's essential features. His approach appears to define philosophy in terms of one particular methodological preference without explaining why this is justified. Some other consequences extend beyond excluding historical figures. Some contemporary work employs case studies, "intuition pumps", and narrative scenarios that do not fit Russell's model of explicit argumentation. David Chalmers' p-zombie thought experiment heavily relies on conceivability intuitions. His arguments against physicalism are more extended, but the construction of that thought experiment as a conceivable scenario is, according to him, sufficient for it to be possible. So, this move from the conceivability of the scenario to its possibility depends on intuition, rather than argumentation, and this by itself is enough to undermine the physicalist solution for the hard problem of consciousness.

Murray Smith also attempts a restrictive approach by laying out the specific normative criteria for philosophy, as we have seen, i.e., rational warrant, empirical support and reflective maturity (Smith 2016, 187-89). He argues that these norms make philosophy "more demanding than science and significantly more challenging than art in terms of the criteria of rationality,

empiricism, and reflexivity." This contrasts sharply with art's primary objective, which he describes as delivering "an adventure in perception, cognition, and emotion" where various values "interpenetrate one another" (Smith 2006, 185).

While Smith's approach appears more explicit in its metaphilosophical analysis and more complex than Russell's methodological essentialism, it faces similar exclusionary problems. The requirements for rational warrant, as analysis of concepts underlying our beliefs, or for empirical support, would disqualify significant areas of philosophical work. Much metaphysics, logical theory, and normative ethics proceed through purely *a priori* reasoning. Chalmers' argument about p-zombies is such an example. Smith's requirement for "reflective maturity" also proves problematic when examined closely. While the ideal of examining beliefs from a multitude of angles, with a "reflective movement back and forth, between examples and principles and between competing theories" (Smith 2016) sounds compelling, it is unclear how this standard would be applied consistently in every philosophical work. Many philosophical works pursue a particular perspective with a singular focus rather than attempting broad analysis and back and forth. In cases like these, the most we could do is to accuse such work of poorer quality rather than outright exclude it from philosophy. Smith's normative approach also faces the problem of arbitrary boundary-drawing. He acknowledges that there may be "an area of overlap between the domains of art and philosophy," citing documentaries like John Berger's *Ways of Seeing* as potentially philosophical works. However, he provides no principled way to distinguish when artistic works cross into philosophical territory.

Russell's and Smith's approaches fail the first criterion for evaluating metaphilosophical positions. Russell's insistence on explicit argumentation would imply that some philosophers, as we have seen, and many others have been "doing philosophy in completely the wrong way" (Overgaard et al. 2013, 25) because they employed literary devices, ordinary language examples, and intuitive appeals rather than formal arguments.

7.1.2 Inclusive metaphilosophies

While restrictive metaphilosophies like Russell's and Smith's exclude too much recognised philosophical work, the opposite problem emerges with overly inclusive approaches that risk making philosophy indistinguishable from any form of thinking. These approaches face the challenge of maintaining philosophy's distinctive identity while remaining open to methodological innovation.

Stephen Mulhall's approach to film as philosophy, inspired by Stanley Cavell's understanding of philosophical practice, exemplifies some of the challenges with overly inclusive metaphilosophies. Mulhall's position, while avoiding the restrictive problems I identified in Russell and Smith, seems to open the door so widely that it becomes difficult to maintain any boundaries around philosophical practice.

Drawing from Cavell, Mulhall argues that philosophy should be understood as the "radicalization of questioning." According to this view, philosophy emerges from a natural human tendency to reflect on our practices and beliefs, distinguished only by the persistence and concentration with which philosophers pursue such questioning. It is in this way that the philosopher regards philosophy:

> (1) There is no essential break between natural human reflection and "inveterate reflexivity of philosophy," the distinction being that the philosopher questions more persistently;
>
> (2) philosophy can occur anywhere or at any time – any human who has a certain practice (the examples he gives us are those of a scientist or a literary critic) may question the nature and resources of it, and the same kind of radical questioning (or philosophy) can be made by a filmmaker;
>
> (3) Since philosophy questions and reflects on other human activities, it must do the same about itself (Mulhall 2005, 29).

Mulhall quotes Cavell on this idea when he says that philosophy is "a willingness not to think about something other than what ordinary human beings think about" (Cavell cited by Mulhall 2008, 144). It is a fact that every human, at some point, thinks about the morality of specific actions, the justice of certain principles, the meaning of their existence, the nature of art, and so on. These are the "big", "eternal", philosophical questions.

This conception has appealing features. It avoids the arbitrary exclusions that affect essentialist approaches, recognises philosophy's continuity with everyday reflection, and acknowledges its self-reflexive character. Mulhall's insistence that films can teach us "what philosophy might be" rather than requiring films to conform to predetermined philosophical categories represents an open and democratic attitude.

However, it seems that this would imply that every human work can be philosophy, which is not very helpful when we try to decide whether a film philosophises or not. Thus, such an inclusiveness creates some difficulties. First, it makes it nearly impossible to distinguish philosophy from other intellectual disciplines; methodological distinctions can be useful for that. Second,

Mulhall's approach provides no clear guidelines for evaluation. If any persistent questioning potentially constitutes philosophy, how do we distinguish good philosophical work from poor philosophical work? A complete approach should be able to respond to this, as explored further in what follows.

We can see that excessively inclusive conceptions of philosophy suffer from similar problems to restrictive conceptions. If, on the one hand, exclusive conceptions reduce the nature of philosophy to specific methods, on the other hand, inclusive conceptions just centre on its topics, which "launches too wide a net" (Overgaard et al. 2013, 22).

7.2 Alternatives to essentialist metaphilosophies

The restrictive and inclusive metaphilosophies examined thus far work with assumptions about the fundamental legitimacy of traditional philosophical boundaries, even when they disagree about where precisely to draw them. However, more radical challenges to the film as philosophy debate emerge. We can challenge how such conceptions of philosophy may systematically exclude certain voices and ways of knowing, thereby reproducing epistemic injustices.

Miranda Fricker's account of epistemic injustice may offer an insight into how the film as philosophy debate itself may instantiate forms of exclusion. Fricker identifies two primary forms of epistemic injustice: testimonial injustice, which "occurs when prejudice causes a hearer to give a deflated level of credibility to a speaker's word", that is "caused by prejudice in the economy of credibility"; and hermeneutical injustice which "occurs at a prior stage, when a gap in collective interpretive resources puts someone at an unfair disadvantage when it comes to making sense of their social experiences", "caused by structural prejudice in the economy of collective hermeneutical resources" (Fricker 2007, 1).

Both forms of injustice operate through the "social imagination" which is "a collective bundle of concepts, ideas, and stereotypes that provide the background assumptions and paradigms within which epistemic interactions take place" (Cunliffe 2019, 170).

When applied to the film as philosophy debate, this reveals how the terms of the discussion may themselves embody epistemic injustice. The restrictive approaches do not merely establish methodological criteria; they may actively exclude ways of knowing that have historically been marginalised within academic philosophy. An insistence on explicit argumentation, for instance, privileges a particular mode of rational discourse that has been coded as aggressive, competitive

and masculine within philosophical tradition, while dismissing more exploratory, associative, or imagistic approaches to philosophical thinking (Rooney 2022).

As Catherine Constable argues, the crisis in film theory reflects broader tensions about "the status and value of theory and theorising" in which "complexity is not an indication of lack of efficacy and the claim that all theory should be simple if it is to be useful is itself a problematic assumption" (Constable 2005, 2). Constable's work demonstrates how feminist philosophers like Michèle Le Doeuff have challenged the traditional separation between theory and everyday experience, arguing instead that "there is no such thing as 'theory per se' because all theories have practical implications, constructing the world and the place of the people in it in different ways" (Constable 2005, 2). This perspective reveals how apparently neutral methodological criteria may actually serve to maintain existing power relations within the academy.

The hermeneutical dimension of epistemic injustice is also relevant to understanding the exclusions built into traditional philosophical discourse. As Constable notes, feminist philosophy "has repeatedly challenged traditional conceptions of the nature of theory, offering diverse accounts of the future of theorising" (Constable 2005, 3). Yet these alternative conceptions often struggle to gain recognition within philosophical institutions precisely because the dominant hermeneutical resources are shaped by the exclusion of feminist and other marginalised perspectives.

This suggests that the film as philosophy debate requires interrogating how metaphilosophical criteria may embody systematic exclusions. Film's contribution to philosophy may lie not in conforming to traditional philosophical methods, but in providing what Zoe Cunliffe terms "epistemic correctives" that can "put pressure on prejudicial stereotypes and contribute to a broadening of the social imagination" (Cunliffe 2019, 170). Through narrative fiction, film can offer "familiarization" with marginalized perspectives, increase awareness regarding systematic prejudices, and provide "representation for marginalized groups" in ways that challenge the dominance of traditional philosophical approaches (Cunliffe 2019, 171-173).

Furthermore, cinema's capacity to influence the social imagination positions it as particularly suited to addressing hermeneutical injustice. Films can "highlight the inadequacy of existing concepts in the social imagination to capture all lived experiences" and "facilitate the development of an open-minded hermeneutical climate" that enables the generation of alternative meanings (Cunliffe 2019, 174-175). In this case, film as philosophy could represent a fundamental challenge to the epistemic injustices embedded within traditional philosophical discourse.

A more radical account that can challenge essentialist metaphilosophies emerges from François Laruelle's "non-philosophy" and its application to cinema by John Ó Maoilearca (Mullarkey). Rather than arguing for either restrictive or inclusive definitions of philosophy, non-philosophy questions the entire attempt at defining philosophy, exposing what Laruelle calls philosophy's fundamental circularity.

According to Laruelle, philosophy operates under what he terms the "Principle of sufficient philosophy", which is philosophy's "absolute autonomy, its essence as self-positing/donating/naming/deciding/grounding" (Laruelle 1999, 139). This principle guarantees philosophy's authority over other disciplines whilst maintaining the "impossibility for philosophy to constitute a rigorous, non-circular thinking of itself" (Ibid.). Philosophy remains trapped in what Laruelle describes as "self-reflection, self-consciousness; it thinks, or in the best of cases, feels that it thinks when it thinks" (Ibid.). Every philosophical approach to any subject, including film, inevitably "makes a 'circle with its object'" by imposing its own interpretative framework upon that which it claims to study (Mullarkey 2009, 204).

The implications for the film as philosophy debate are important. Whether we adopt restrictive or inclusive approaches, we remain trapped within what Laruelle identifies as philosophy's essential structure: "What philosophy calls the reality of the subject is always its concept of the subject" (Ibid.). Any attempt to determine whether films can philosophise requires "a transcendental stance, or prior model of philosophy" that inevitably "infects" cinema "with that philosophy's own nature" (Ibid.). The debate itself presupposes precisely what it seeks to establish.

Mullarkey applies this directly to cinema, arguing that "the non-philosophical Real of film has rebounded each and every philosophical attempt to explain it exhaustively" (Ibid., 205). Rather than offering another definition of what film is, he proposes "a thinking alongside the Real" that maintains "passivity towards film in its thoughtful resistance to extant philosophy" (Ibid.). This approach abandons the search for necessary and sufficient conditions of philosophy to apply to film.

Central to this is the concept of an "élan cinématique" which represents "the force of which undoes any attempt at saying what film is – world-view, concept creator, cultural product, unconscious identification, information process" (Mullarkey 2009, 215). From this perspective, film "thinks, but in a non-philosophical way" by "resisting what philosophy thinks film and so also it itself are when philosophy applies itself to film and illustrates itself through films" (Ibid.).

This resistance can also derive from cinema's hybridity, as it is "part theatrical, part literary, part photographic, part musical, a group-manufactured commercial art-product", and this makes it "art which ostensibly resists definition like few others" (Mullarkey 2009, 215). Thus, the "very messiness of film" merges with "the same messiness of reality" in ways that frustrate philosophical systematisation (Ibid.).

From a non-philosophical perspective, instead of asking whether films can philosophise, we should examine how cinema's resistance to theory reveals something fundamental about the nature of both philosophy and reality. As Mullarkey suggests, cinema "unphilosophizes what philosophy – and thinking – thought it was" (Mullarkey 2009, 215).

This non-philosophical account offers a distinctive response to the metaphilosophical tensions I have examined. Unlike restrictive approaches that exclude too much, or inclusive approaches that include too much, non-philosophy suggests that the very framework of inclusion and exclusion may be misguided. The question is not whether we should expand or contract philosophy's boundaries, but whether boundary-drawing itself reflects a philosophical prejudice that distorts our understanding of both cinema and thinking.

However, the non-philosophical approach might also be contested. If philosophy always "makes a circle with its object," non-philosophy might fall into the same circularity itself. Laruelle addresses this with his notion of "unilateral duality": a relationship that is "not one which has two sides"; instead, a relationship where "the Real does not constitute a side, only non-philosophy or philosophy's relative autonomy does so" (Laruelle 1999, 143). However, this may constitute another form of philosophical system-building, while denying its systematicity. Moreover, the practical consequences of non-philosophy for film analysis are unclear. While the critique of philosophical circularity is interesting, it does not offer clear guidance for understanding philosophical meaning in specific films.

We should still consider challenges to essentialist accounts, as they point toward the need for moderate metaphilosophical approaches that avoid both the extremes of essentialism and the abandonment of any philosophical nature or rigour entirely.

7.3 A moderate metaphilosophical approach

It is possible to say that philosophers are not very well equipped to decide if films can philosophise, and what philosophy is in general. As Gary Gutting puts it, the main reason metaphilosophy is not

that interesting is that when philosophers defend their metaphilosophy, they have a "dogmatic attitude that derives the nature of philosophy from controversial philosophical doctrines" and "an abstract, overly generalized approach that pays no attention to the details of philosophical practice" (Gutting 2009, 2). Wartenberg also thinks the question of what philosophy is an "intractable" problem (Wartenberg 2007, 28).

An easy solution would be a deflationary or institutional approach that appeals to what has been recognised as philosophy historically. In this case, philosophy is "whatever the people who are employed *as* philosophers at universities and other institutions do, or whatever material librarians catalog as such – end of story" (Overgaard et al. 2013, 20). This approach could explain what philosophy is and respond to objections against FAP. For example, appealing to existing philosophy that lacks explicit argumentation can be an application of this strategy.

One version of this approach would be the "reader-response" strategy, where "any text [or work] that gets read as philosophy, from Shakespeare to Darwin, counts as philosophy. Philosophy lies in the eye of the beholder" (Blackburn 2004, xiii).

However, as Smith argues, this strategy "drains the debate of part of its lifeblood, leaving us with a rather anemic general solution: whatever someone somewhere has designated 'philosophy' can serve as a measure of whether film can philosophize" which is a "low threshold to meet" (Smith 2019). In this case, if anything can count as philosophy simply by being designated as such, then the FAP hypothesis becomes trivial: films can philosophise whenever anyone chooses to interpret them philosophically. This approach also faces an arbitrariness problem: different institutions, communities, and historical periods recognise different works as philosophical. If philosophy is simply whatever gets institutionally recognised, then philosophical status becomes relative to particular social contexts rather than grounded in any objective features of the works themselves.

The deflationary strategy also faces a lack of explanatory power. If philosophy simply is whatever gets called philosophy, then questions about film's philosophical potential become sociological rather than normative. We can investigate how various communities have responded to films, whether film scholars have successfully established institutional recognition for cinematic philosophy, and what strategies prove effective in expanding philosophical boundaries. But we cannot address the substantive question of whether films actually possess the features that make philosophical contributions valuable.

Inclusive or alternative metaphilosophical accounts are more democratic, as they recognise intellectual work from other disciplines and cultural practices as philosophical. However, this "democratic" appeal tends to collapse accessibility with standards, and inclusive or alternative metaphilosophies may threaten philosophy's ability to maintain a coherent identity. Making philosophy more accessible by connecting it to popular culture, recognising its presence in various human activities, and welcoming contributions from diverse communities is an important goal. But accessibility does not require forsaking all traditional criteria for evaluating the nature of philosophy or the value of a philosophical work. Philosophical progress, to whatever extent it occurs, depends on the discipline's ability to distinguish more successful from less successful methods and approaches to problems – philosophy needs its internal resources for evaluating and improving methods.

There is, thus, a challenge to develop criteria that are open to innovation but still preserve philosophical identity, i.e. ways of distinguishing philosophical from non-philosophical work, but also good from bad philosophical work. A solution might be to view philosophy's boundaries as somewhat flexible to allow exchange and collaboration while maintaining distinctive functions.

As we saw earlier, Wartenberg presented his metaphilosophical account to defend FAP, with three conceptions to be regarded inclusively, mutually supportive: philosophy as approaching a set of "eternal questions" (a topical account); philosophy as a meta-discipline (also a topical account); and philosophy as a set of methods (methodological account) (Wartenberg 2007, 29). This approach seems to be adequate and to have good explanatory power. The main insight of this approach is the concession that these three conceptions are mutually supportive: the fact that a work engages with one dimension may not automatically qualify it as philosophical, but the combination of dimensions may do so.

This approach avoids some problems with essentialist metaphilosophies. Contrary to methodological essentialisms, like Russell's or Smith's, it does not reduce philosophy to a single method, like explicit argumentation, and unlike topical essentialisms, it does not make philosophy indistinguishable from any form of persistent questioning.

Wartenberg's approach can be supplemented with Rescher's theory of orientational pluralism in order to understand the relative progress of philosophy and persistent disagreements.

Rescher argues that philosophical disagreement originates in differences in what he calls "orientational commitments", fundamental differences in how philosophers approach the evaluation of arguments and evidence. These commitments involve "judgments of plausibility,

principles of presumption – intellectual predilections" that shape how philosophers weigh different considerations (Rescher 2006, 129). These orientations concern methods, affecting the "probative methods of inquiry and substantiation" that philosophers employ when evaluating positions (Ibid.).

These orientational differences arise from "our culture, the 'spirit of the times'" (Ibid.). Importantly, these disagreements are about values, not facts (two philosophers may agree on the same evidence present in an argument while reaching different conclusions).

If we integrate Wartenberg's conceptions with Rescher's orientational pluralism, we may be able to approach a better moderate metaphilosophical position. This combined Wartenberg-Rescher approach passes both of Overgaard et al.'s tests for adequate metaphilosophical positions. First, it passes the non-revisionist criterion: it avoids being overly revisionist because it accommodates most recognised philosophical work. Different historical philosophers and contemporary approaches can be understood as engaging with Wartenberg's three dimensions from different orientational perspectives. Second, it explains philosophy's relative lack of progress: Rescher's might explain why philosophy lacks the cumulative progress seen in science: "It is the comparative improminence of conflicting methodological orientations in science – i.e., the consensus regarding cognitive values in the domain – that decisively distinguishes science from philosophy and explains why the development of science is comparably discord-free" (Ibid., 130). Even if, for instance, philosophers would agree on Wartenberg's indication of those three conceptions of philosophy, their different orientational commitments lead them to evaluate philosophical arguments and theories differently, preventing the kind of consensus that enables scientific progress.

This also explains why the FAP debate persists: philosophers bring different probative orientations to evaluating how films engage with eternal questions, function as meta-disciplinary reflection, or employ philosophical methods. The debate reflects legitimate differences in philosophical orientation, making it inherently resistant to a definitive resolution.

This approach, which we might see as a form of pluralistic essentialism, views philosophy as having essential features that work together, and so it has several advantages over both restrictive and overly inclusive alternatives. It "passes" Overgaard et al's test, as we have seen. It is also flexible enough to accommodate philosophical diversity without abandoning all standards and allowing for methodological innovation. It also explains why borderline cases generate deep disagreement: when we debate whether a particular film counts as philosophy, we are usually weighing how well it satisfies different dimensions of philosophical practice. Maybe some works

satisfy all dimensions, others may satisfy none. Furthermore, it may provide resources for evaluating quality within philosophical work. Even if we agree that a film satisfies the criteria for philosophical status, we can still ask how well it addresses fundamental questions, how well it employs philosophical methods, and how insightfully it reflects on disciplinary boundaries, like its own nature. Finally, it acknowledges philosophy's historical development, maintaining continuity with tradition. New methodological approaches can qualify as philosophical if they engage with traditional philosophical questions and contribute to ongoing philosophical conversations. This allows for innovation without requiring complete breaks with philosophical history.

The moderate position I am adopting embraces democratic openness without collapsing philosophy's identity. It welcomes contributions from diverse sources and recognises that valuable philosophical work can occur outside traditional academic boundaries, while maintaining standards for distinguishing philosophical from non-philosophical contributions. This is evident in several features of the approach.

1) It does not privilege any particular methodological approach as essentially philosophical. Traditional argumentation, thought experimentation, conceptual analysis, phenomenological description, and other methods can all count as philosophical if employed in appropriate contexts. This openness allows for the kind of methodological diversity that has characterised philosophy throughout its history;
2) It does not restrict philosophical authority to professional philosophers. Filmmakers, scientists, artists, and others can potentially make philosophical contributions if their work engages appropriately with philosophical tradition. This recognition reflects the historical reality that many important philosophical insights have come from individuals whose primary identity was not as professional philosophers.
3) It acknowledges that philosophy benefits from engagement with other disciplines and cultural practices. Rather than treating philosophy as an isolated academic speciality, this moderate approach recognises that philosophical questions arise throughout human culture and that philosophy's vitality depends partly on its ability to engage with broader cultural conversations.

However, this approach is constrained. Not every form of thinking counts as philosophical, and not every engagement with philosophical themes constitutes philosophical contribution.

This balance between openness and constraint maintains philosophical standards not through exclusionary gatekeeping but through evaluative criteria that can accommodate revision.

The pluralistic essentialist framework creates space for methodological innovation while preserving philosophy's capacity for self-correction. When films like those in my case studies demonstrate unexpected ways of engaging with philosophical questions, we can recognise their contributions without abandoning resources necessary for distinguishing more from less successful philosophical work. This enables philosophy to learn from its encounters with cinema while maintaining its identity.

7.3.1 Moderate pro-film as philosophy

The moderate approach adopted here meshes with the moderate pro-cinematic position defended by Wartenberg, which "focuses upon specific techniques that filmmakers can employ to do philosophy in film, most centrally the thought experiment" (Wartenberg 2016).

An important feature of this approach is its emphasis on particular analysis rather than general theory. Wartenberg argues that questions about whether specific films philosophise should be addressed case-by-case through examination of particular works rather than through a priori theoretical arguments (Wartenberg 2007). Despite his motivation of understanding whether film can serve as a medium for philosophising, Wartenberg does not leap ahead, beginning instead with a sceptical attitude toward this possibility. Thus, he does not argue or imply that all films can philosophise; whether a film philosophises is a matter of close analysis of how exactly the film philosophises and about what. So, if on the one hand we should not presuppose that films, in general, philosophise, or are in general better equipped to do so for any specific reason, on the other, we should not presuppose a priori that no film is capable of doing philosophy either, as Mulhall argues (Mulhall 2008, 131-32). Wartenberg agrees with Mulhall on this, rejecting a priori arguments about cinema's inability to philosophise (Wartenberg 2007, 14)

This emphasis on particular analysis does not abandon theoretical reflection entirely. Rather, it suggests that theoretical claims about film's philosophical potential should emerge from and be tested against careful examination of specific cases. The relationship between theory and practice becomes dialectical: theory guides particular analyses, while particular analyses refine and sometimes revise theory. My working definition of philosophy, combined with my account of thought experiments and cinematic techniques, provides tools that help identify what philosophical dimensions might be present in specific films without predetermining what we will find. It helps us recognise philosophical work when it occurs while remaining sensitive to the distinctive ways cinema might contribute to philosophical understanding. The case-by-case approach thus becomes

justified: since films can engage with my working definition in diverse ways and to varying degrees, only careful attention to particular works can reveal how cinematic techniques generate philosophical insights.

So, the moderate pro-film as philosophy adopted here includes a pluralistic essentialist account of philosophy; a preference for particularist approaches that focus on case-by-case analyses of particular films, the methods they employ or embody (methodological conception), and what philosophical topics they address (topical conception); and a complementary thesis that regards different media as possibly being able to make distinctive but complementary contributions to philosophical understanding. Thus, with this approach we refocus the question "are films able to philosophise?" to "how can films contribute to philosophical understanding, alongside traditional philosophy?" The FAP debate and how it was defended in this thesis may apply to other intellectual and cultural activities and art forms like literature, science, or music; however, I remain agnostic on whether these other practices might be comparatively better or worse equipped for philosophical practice.

We should consider whether we have been approaching this entire question wrongly. Throughout this thesis, we have examined philosophy through Wartenberg's three conceptions: as addressing eternal questions, as a meta-discipline, and as a set of methods. We have debated whether films can meet these criteria. But what if these debates miss something fundamental about what philosophy actually is and why it matters?

Throughout this thesis, we have explored whether films can do philosophy through thought experiments. But we should also ask: what is the point of doing philosophy at all? This question matters because if philosophy itself lacks a clear purpose, then arguing for film's philosophical potential becomes meaningless.

Ingmar Persson argues that "the fundamental problem of philosophy is whether doing it has any point, since if it does not have any point, there is no reason to do it" (Persson 2018, 52). He suggests that philosophy's main goal should be "to establish a rational consensus about what the answers to its main questions are" (Ibid., 52). But philosophy seems unable to achieve this. Philosophical arguments are "bound to be inconclusive" (Ibid., 52), leaving us without the clear answers we seek.

This might seem to undermine philosophy entirely. But Persson offers a different way to think about philosophy's value. Even if philosophy cannot solve its big questions definitively, it can

still serve important purposes. Philosophy "generates an increasing number of finer grained distinctions in terms of which we try to conceptualize reality, and this is a sort of progress" (Ibid., 52). We get better at understanding the complexity of problems, even if we cannot solve them completely.

More importantly, philosophy can help us know ourselves and guide our lives. Persson argues that when philosophical arguments cannot decide between different positions, "our personalities might slip in to do so. Our philosophy will then express our personality" (Ibid., 52). This is not a failure of philosophy but a feature. Philosophy becomes a way of working out who we are and what we believe.

Philosophy can also provide us with "something by which we can live" when our conclusions "have practical import" (Ibid., 52). Ancient philosophers understood this. As Persson notes, according to Epicurus, "Philosophy is an activity that secures the flourishing [eudaimon] life by arguments and reasoning" (Ibid., 62). Philosophy was not just about abstract thinking but about learning how to live well.

This view of philosophy supports our argument for film as philosophy. If philosophy's purpose includes self-knowledge and life guidance, then films may be particularly well-suited to contribute. Films show us human situations in rich detail. They engage our emotions alongside our reasoning. They present us with scenarios that help us think through our own values and choices. Persson emphasises that philosophy works best when it "combines the theoretical and practical: a philosophical vision of the world encompassing practical implications to live by" (Ibid., 68). This is exactly what many philosophical films do. They do not just present abstract arguments but show us those arguments working through concrete situations involving real human concerns.

The point of philosophy, then, is not just to succeed at arguments or establish universal truths. Philosophy helps us understand ourselves, guides our choices, and provides us with ways of thinking about our lives. Films can contribute to these goals in ways that complement traditional philosophical texts. Through thought experiments presented in cinematic form, films can engage our rational and emotional minds, helping us work through philosophical problems in ways that matter for how we actually live.

This reframing changes everything about the FAP debate. Instead of asking whether films can meet philosophy's methodological requirements, we should ask whether they can serve philosophy's deeper purposes. Instead of worrying about explicit argumentation or thought experimentation, we could consider whether films help us understand ourselves and our world

better. Instead of demanding universal truths, we could value the rich distinctions and insights that films can offer.

When we see philosophy this way, as a practice of self-discovery, life guidance, and conceptual refinement, film's contributions become not just acceptable but essential. Films excel at exactly what Persson suggests philosophy should do: they help us work through who we are, they guide us in thinking about how to live, and they offer us new ways of understanding complex human situations.

The technical debates we have explored throughout this thesis about thought experiments, medium specificity, and authorial intention remain important. But they are not the final word. What matters most is whether films can contribute to the kind of philosophical understanding that helps us live better, think more clearly, and know ourselves more deeply. On this measure, the case for film as philosophy is not just defensible: it is compelling.

If this broader understanding of philosophy's purpose is all we need, then the question is resolved. Films can generally help us understand ourselves, guide our choices, and refine our thinking about complex human situations. But if we want a more focused discussion of philosophical methods specifically, then the work we have done throughout this thesis becomes essential. At the very least, we can recognise that thought experiments are important tools for reconceptualising our assumptions, and films can certainly do that work too, as was shown.

This chapter has explored the metaphilosophical foundations that underlie the film as philosophy debate. As we have seen throughout this investigation, the arguments for and against FAP rest on specific assumptions about what philosophy is and how we should conduct it. We cannot resolve this debate by examining cinema's capacities alone, we must also confront fundamental questions about the nature of philosophy itself.

We began by examining how restrictive metaphilosophies create artificial barriers to film's philosophical potential. The explicitness and generality objections, as defended by Russell and Smith, exemplify essentialist approaches that define philosophy through supposedly necessary and sufficient conditions. Russell's insistence on explicit argumentation and Smith's emphasis on rational warrant, empirical support, and reflective maturity both exclude too much recognised philosophical work. As I demonstrated, these approaches fail the non-revisionist criterion by suggesting that many historical and contemporary philosophers have been "doing philosophy in completely the wrong way."

We then considered overly inclusive approaches that risk making philosophy indistinguishable from any form of persistent questioning. While Mulhall's Cavellian conception of philosophy as "radicalised questioning" avoids the exclusions of restrictive approaches, it provides no clear guidelines for distinguishing philosophical from non-philosophical work, or for evaluating the quality of philosophical contributions.

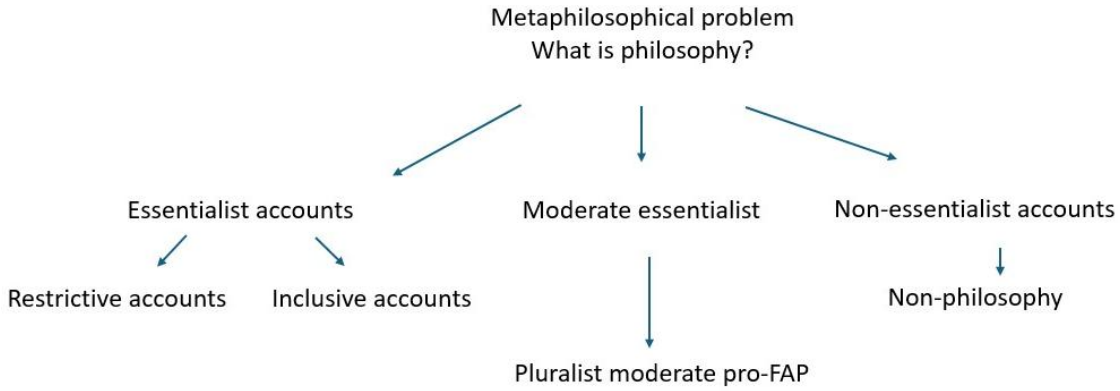
We also examined alternative challenges to traditional metaphilosophical approaches. The analysis of epistemic injustice reveals how apparently neutral methodological criteria may systematically exclude marginalised voices and ways of knowing. Laruelle's non-philosophical critique exposes philosophy's fundamental circularity: its inability to constitute rigorous, non-circular thinking about itself. While these alternatives highlight important problems with essentialist approaches, they offer limited practical guidance for evaluating specific cases of cinematic philosophy.

In response to these challenges, I defended a moderate metaphilosophical approach that combines Wartenberg's three mutually supportive conceptions of philosophy with Rescher's orientational pluralism. This approach views philosophy as engaging with eternal questions, functioning as a meta-discipline, and employing distinctive methods. We can understand persistent disagreements in the FAP debate as reflecting differences in philosophical orientation.

This moderate position enables me to adopt a moderate pro-film as philosophy approach. Rather than defending or rejecting the philosophical potential of cinema in general, I focus on case-by-case analyses of particular films. We ask not whether films can philosophise, but how specific films contribute to philosophical understanding alongside traditional philosophical methods. This approach maintains standards for distinguishing philosophical from non-philosophical contributions while remaining open to methodological innovation and diverse voices and means of expression.

The metaphilosophical framework I developed allows us to evaluate films as potential philosophical thought experiments without falling into either restrictive essentialism or unlimited inclusiveness. As demonstrated earlier through my case studies, films can engage with philosophical questions, employ philosophical methods, and reflect on the boundaries of disciplines, including their own nature as cinematic works. The key insight is that we should approach each film individually, examining how it employs cinematic techniques to contribute to ongoing philosophical conversations.

Diagram 2: Metaphilosophical Approaches to Film as Philosophy





Conclusion

This thesis has developed and defended a revised bold thesis about cinematic philosophy: films can produce philosophy through thought experimentation, and they do so in ways that are characteristically cinematic rather than identical to traditional philosophical practice. Through an investigation about the film as philosophy debate, thought experiments, Wes Craven's films and metaphilosophy, I have established that the boundaries between philosophical work and popular culture may need reconsidering.

My argument proceeded through three interconnected claims that proved challenging to develop.

The first challenge was demonstrating that philosophical thinking does not always require explicit argumentation. Many traditional objections to film as philosophy rest on overly restrictive conceptions of philosophical method that privilege explicit argumentation.

Working through the extensive debates surrounding thought experiments between John Norton, James Robert Brown, and various moderate inflationist positions was essential for showing that philosophical thinking can occur through methods that do not conform to such narrow constraints. Norton's argument view attempts to reduce all thought experiments to straightforward arguments, claiming they possess no distinctive epistemic value beyond what explicit argumentation provides. Brown's Platonic view goes to the opposite extreme, treating thought experiments as providing quasi-perceptual access to abstract Platonic laws of nature. Both

positions miss something important about how thought experiments actually function in philosophy.

The breakthrough came through examining moderate inflationist positions that avoid both extremes. Philosophers like Tamar Gendler, Nancy Nersessian, and Nenad Miščević have developed accounts of how thought experiments generate knowledge through mental modelling, constructivist processes, and quasi-observational methods. These accounts reveal that thought experiments operate through cognitive mechanisms that are both more complex than simple argumentation and more naturalistic than Platonic intuition.

This required diving into philosophy of science literature. The challenge was connecting these epistemological theories to concrete questions about cinematic techniques. But this effort revealed that thought experiments, and by extension, cinematic thought experiments, can generate philosophical understanding through processes that engage our cognitive and emotional resources in ways that purely abstract argumentation cannot.

The second major challenge was demonstrating that films offer distinctive advantages as philosophical thought experiments. This required more than showing that cinema can do what traditional philosophy already does; it demanded establishing that cinematic techniques provide unique resources for philosophical investigation that textual methods cannot match.

I adopted and extended David Davies and Tom McClelland's strategy of focusing on cinema's specific capacities rather than attempting to defend film as philosophy in general. Films possess unique capacities through their ability to present rich, detailed scenarios that engage viewers' cognitive and emotional resources in ways that textual thought experiments cannot. Where traditional philosophical thought experiments often rely on sparse descriptions that leave critical details unspecified, cinema can construct fully realised scenarios that activate multiple cognitive systems simultaneously.

Drawing on mental modelling accounts developed by philosophers like Tamar Gendler, Nenad Miščević, and Nancy Nersessian, I demonstrated how films could activate tacit knowledge that remains inaccessible to purely abstract reasoning. Cinema's audiovisual presentation enables what Gendler calls "quasi-observational" processes that allow viewers to track features of imaginary scenarios through perceptual-like engagement. This addresses longstanding concerns about thought experiments' reliance on inadequately specified background conditions, a criticism developed by Kathleen Wilkes.

I was able to connect these theories about mental modelling to concrete analysis of how specific cinematic techniques generate philosophical understanding. Editing creates temporal relationships that can model causation and logical connection; framing establishes perspective and directs attention to philosophically relevant features; sound design shapes interpretation; mise-en-scène constructs detailed worlds that enable quasi-observational engagement with philosophical scenarios.

The third challenge was resolving the imposition objection. This objection holds that any philosophical content we find in films is projected by interpreters rather than present in the films themselves. If this objection succeeds, claims about cinematic philosophy fail. Traditional approaches to authorial intention seemed to lead either to rigid author-centrism that ignored cinema's collaborative nature or to unlimited reader-response approaches that dissolved any constraints on interpretation.

The solution emerged through studying Paisley Livingston and Vítor Moura's approaches to interpretation, particularly Moura's concept of the "spectral author", a presence that emerges from the work itself while retaining connection to the actual director. Instead of treating films as determining fixed messages from directors, we can understand them as establishing conditions for collaborative meaning-making.

This enables a collaborative philosophical reasoning model. Filmmakers establish conditions for philosophical investigation through cinematic techniques, creating scenarios, directing attention, and providing evidence, while viewers participate actively in developing philosophical implications through interpretation that remains grounded in evidence from the films themselves. The philosophical work emerges through this collaboration rather than being entirely determined by either filmmaker's intention or the viewer's interpretation.

My detailed analysis of Wes Craven's films provided empirical validation for these theoretical bases. I chose Craven partly because of his formal philosophical training, but mainly because his work consistently addresses fundamental philosophical questions in systematic ways.

Each film demonstrated different types of philosophical thought experimentation in action.

The Last House on the Left (1972) functions as a powerful counterexample to conventional moral intuitions about revenge and justice. The film constructs a scenario that tests common assumptions about moral symmetry (the idea that equivalent wrongs justify equivalent responses). Through its brutal, realistic style and careful attention to the psychological effects of violence, the film forces viewers to confront uncomfortable questions about whether revenge can be morally

justified. Craven uses parallel editing to establish equivalences between the criminals' violence and the parents of victim's revenge, suggesting moral symmetry while simultaneously showing the psychological costs of both. The film's realistic style refuses to aestheticise violence or provide comfortable moral distance, forcing viewers to confront the actual implications of revenge. The film does not argue explicitly against revenge, but creates conditions that enable viewers to discover problems with conventional revenge justifications.

A Nightmare on Elm Street (1984) demonstrates cinema's capacity for possibility-establishment, another type of philosophical thought experiment. The film explores philosophical questions about consciousness, reality, and the relationship between dreaming and waking experience by creating a scenario where the boundaries between dream and reality become unclear.

The film's philosophical investigation is conducted through its manipulation of cinematic techniques for representing consciousness. Craven uses editing, sound design, and visual effects to create uncertainty about characters' mental states and the ontological status of their experiences. The film's careful attention to the phenomenology of dreaming (the way dream logic works, how dreams incorporate elements from waking experience, how consciousness transitions between states) provides a detailed exploration of philosophical questions about the nature of consciousness and reality. The film establishes the possibility that our ordinary criteria for distinguishing reality from illusion might be inadequate. It does not argue that external reality does not exist, but demonstrates that subjective experience alone may be insufficient for determining the nature of reality.

Scream (1996) presents as metacinematic commentary that investigates philosophical questions about media effects, hyperreality, and the relationship between fictional and real violence. The film's self-referential structure enables examination of how media narratives shape behaviour and expectations. The film's philosophical work occurs through its investigation of the relationship between horror film conventions and real-world behaviour. Characters' explicit knowledge of horror film "rules" affects their actions and survival strategies, creating a scenario that models questions about media influence and the relationship between fictional and real experience.

Wes Craven's New Nightmare (1994) demonstrates cinema's capacity for philosophical theory construction through its exploration of fictional ontology and the relationship between artistic creation and reality. The film creates a complex metacinematic scenario where the creation of

horror films apparently enables fictional entities to gain autonomous existence, providing a basis for investigating philosophical questions about the nature of fictional objects, artistic creation, and the relationship between imagination and reality.

The investigation ultimately revealed that the film as philosophy debate cannot be resolved by examining cinema's capacities alone: it requires confronting fundamental questions about the nature of philosophy itself. I examined how restrictive metaphilosophies create barriers to film's philosophical potential, often defending essentialist accounts that exclude too much recognised philosophical work. Russell's insistence on explicit argumentation exemplifies methodological essentialism, the view that philosophy is essentially defined by its use of specific methods like formal argumentation. Smith's emphasis on rational warrant, empirical support, and reflective maturity reflects a similar approach.

At the same time, overly inclusive approaches risk making philosophy indistinguishable from any form of persistent questioning or reflective thinking. Stephen Mulhall's approach to film as philosophy, inspired by Stanley Cavell's understanding of philosophical practice, exemplifies some of these challenges. Drawing from Cavell, Mulhall argues that philosophy should be understood as the "radicalisation of questioning", distinguishable from ordinary reflection only by the persistence with which philosophers pursue such questioning. While this approach avoids the exclusionary problems of restrictive metaphilosophies, it provides no clear guidelines for distinguishing philosophical from non-philosophical work, or for evaluating the quality of philosophical contributions. If philosophy is simply persistent questioning about fundamental issues, then virtually any reflective engagement with basic human concerns could count as philosophical work.

My solution was to defend a moderate metaphilosophical position based on Thomas Wartenberg's three mutually supportive conceptions of philosophy combined with Nicholas Rescher's orientational pluralism. Wartenberg identifies three conceptions that should be regarded inclusively and as mutually supportive: philosophy as approaching a set of "eternal questions" (a topical account), philosophy as a meta-discipline that reflects on the methods and assumptions of other disciplines (also topical), and philosophy as employing a distinctive set of methods (methodological account). This was strengthened by integrating Rescher's theory of orientational pluralism, which provides insight into why philosophical disagreements persist. By integrating Wartenberg's conceptions with Rescher's orientational pluralism, we get what I call a pluralistic

essentialist approach to philosophy. This approach recognises that philosophy possesses essential features (engagement with eternal questions, meta-disciplinary reflection, distinctive methods) that work together, while acknowledging that legitimate differences in orientational commitments lead philosophers to evaluate these features differently, which also explains the relative lack of progress in philosophy and the continuing controversy surrounding film as philosophy.

The most immediate extension of this work would involve a systematic analysis of philosophical thought experimentation in other genres and media forms. Each presents distinctive opportunities and challenges that would test and refine the theoretical grounding I developed.

Science fiction offers particularly promising territory for cinematic philosophy, given the genre's longstanding engagement with philosophical questions. However, I am not sure if the basis I developed for Craven's horror films necessarily transfers directly to science fiction, which works through different genre conventions and concerns. Documentary filmmaking presents another important area for investigation, given its different relationship to truth claims and capacity to engage with real-world philosophical questions through direct examination of actual cases. While my analysis focused on fictional films that create hypothetical scenarios for philosophical investigation, documentaries examine actual people, events, and social conditions that raise philosophical questions. These differences suggest that documentary philosophy may function through different methods than fictional cinematic philosophy, which requires separate consideration.

Research examining filmmakers from different cultural backgrounds could also significantly expand understanding of cinema's philosophical potential. European art cinema directors like Ingmar Bergman, Andrei Tarkovsky, and Jean-Luc Godard have long been recognised for their philosophical engagement, but it could be interesting to analyse how their work functions as philosophical thought experiments. The moderate intentionalist approach here adopted may need to address films created within different cultural contexts and philosophical traditions, particularly those that do not share Western assumptions.

Furthermore, the frameworks I developed for analysing cinematic philosophy may extend to other visual and narrative media that are becoming increasingly important in contemporary culture. Video games, for instance, create interactive narrative experiences that may be interesting to explore. Virtual and augmented reality technologies create new possibilities for philosophical investigation by enabling immersive experiences that approximate the quasi-observational

engagement that I identified as central to cinematic philosophy. Social media and internet culture create collaborative meaning-making practices that work through different methods than traditional media, potentially enabling new forms of collective philosophical investigation.

Throughout this thesis, we have explored whether films can do philosophy through thought experiments. But we should also ask: what is the point of doing philosophy at all? This question matters because if philosophy itself lacks a clear purpose, then arguing for film's philosophical potential becomes meaningless. Philosophy helps us understand ourselves, guide our choices, and refine our thinking about complex human situations. It challenges assumptions and opens new possibilities for thought. If this broader understanding of philosophy's purpose is what matters, then films certainly contribute to these goals. But more than that, they do so in ways that reach audiences philosophy has traditionally struggled to engage. Cinema's integration of intellectual and emotional appeal, its accessibility to diverse audiences, and its cultural prominence position it to extend philosophical thinking into domains that academic philosophy alone cannot reach.

The future of philosophical inquiry may depend partly on its capacity to recognise and cultivate forms of philosophical thinking that operate through means different from those traditionally defining academic practice. As Wartenberg notes,

> One means to enhance the status of philosophy is to show that it is engaged with vital realms of popular culture. While many may see this as watering down philosophy, I think that would be a mistake. To ensure its own vitality, philosophy needs to show its relevance to contemporary people and societies. It is my bet – one on which I have staked a good portion of my life and career – that the linking of philosophy and film is one way to do so" (Wartenberg 2007, 142).

This thesis provides foundations for pursuing this linking while maintaining philosophical standards, though whether we achieve that balance remains an open question.

References

Film and Philosophy Literature

Andersen, Nathan. 2003. "Is Film the Alien Other to Philosophy?: Philosophy as Film in Mulhall's *On Film*." *Film-Philosophy* 7 (2). https://doi.org/10.3366/film.2003.0023.

Arnheim, Rudolf. 1957. *Film as Art*. Berkeley and Los Angeles: University of California Press.

Baggini, Julian. 2003. "Alien Ways of Thinking: Mulhall's *On Film*." *Film-Philosophy* 7 (2).

Bazin, André. 1960. "The Ontology of the Photographic Image." *Film Quarterly* 13 (4): 4-9.

Benko, Steven, and Andrew Pavelich, eds. 2019. *The Good Place and Philosophy*. Chicago: Open Court.

Bordwell, David, and Noël Carroll, eds. 1996. *Post-Theory: Reconstructing Film Studies*. Madison: University of Wisconsin Press.

Bordwell, David, Kristin Thompson, and Jeff Smith. 2017. *Film Art: An Introduction*. 11th ed. New York: McGraw-Hill Education.

Bunce, Robin, and John McCrossin, eds. 2019. *Blade Runner 2049 and Philosophy*. Chicago: Open Court.

Carroll, Noël. 1990. *The Philosophy of Horror: Or, Paradoxes of the Heart.* New York: Routledge.

___. 1996. *Theorizing the Moving Image*. Cambridge: Cambridge University Press.

___. 1999a. "Film, Emotion, and Genre." In *Passionate Views: Film, Cognition, and Emotion*, edited by Carl Plantinga and Greg M. Smith, 21-47. Baltimore: Johns Hopkins University Press.

___. 1999b. *A Philosophy of Mass Art*. Oxford: Oxford University Press.

___. 2006. "Philosophizing Through the Moving Image: The Case of *Serene Velocity*." *Journal of Aesthetics and Art Criticism* 64 (1): 173-185.

___. 2008. "Philosophy in the Moving Image: Response to Bruce Russell." *Film and Philosophy* 12: 17-26.

___. 2013. *Minerva's Night Out: Philosophy, Pop Culture, and Moving Pictures*. Malden: Blackwell.

___. 2017. "Movie-Made Philosophy." In *Film as Philosophy*, edited by Bernd Berzogenrath, 265-285. Minneapolis: University of Minnesota Press.

Carroll, Noël, Laura T. Di Summa, and Shawn Loht, eds. 2019. *The Palgrave Handbook of the Philosophy of Film and Motion Pictures.* New York: Palgrave Macmillan.

Cavell, Stanley. 1996. *Contesting Tears: The Hollywood Melodrama of the Unknown Woman*. Chicago: University of Chicago Press.

___. 1981. *Pursuits of Happiness: The Hollywood Comedy of Remarriage*. Cambridge: Harvard University Press.

___. 1984. *Themes Out of School: Effects and Causes*. Chicago: The University of Chicago Press.

___. 1971. *The World Viewed: Reflections on the Ontology of Film*. Cambridge: Harvard University Press.

Clover, Carol J. 1993. *Men, Women, and Chain Saws: Gender in the Modern Horror Film*. Princeton: Princeton University Press.

Choi, Jinhee. 2006. "Apperception on Display: Structural Films and Philosophy." *Journal of Aesthetics and Art Criticism* 64 (1): 165-172.

Cox, Damian, and Michael P. Levine. 2012. *Thinking Through Film: Doing Philosophy, Watching Movies*. Malden: Wiley-Blackwell.

Danto, Arthur C. 1979. "Moving Pictures." *Quarterly Review of Film Studies* 4 (1): 1-21.

Davies, David. 2007. "Thought Experiments and Fictional Narratives." *Croatian Journal of Philosophy* 7 (1): 29–45.

___. 2008. "Can film be a philosophical medium?" *Postgraduate Journal of Aesthetics* 5, no. 2.

___.2013. "Can Philosophical Thought Experiments Be 'Screened'?" In *Thought Experiments in Philosophy, Science, and the Arts*, edited by Mélanie Frappier, Letitia Meynell, and James Robert Brown, 223-238. New York: Routledge.

___. 2019. "Philosophical Dimensions of Cinematic Experience." In *Philosophy and Film: Bridging Divides*, edited by Christina Rawls, Diana Neiva, and Steven S. Gouveia, 135-156. New York: Routledge.

___. 2019. *Philosophy of the Performing Arts*. Oxford: Oxford University Press.

Davinia, Alice. 2006. "The Rules of Horror: Genre, Representation and Social Context in Contemporary American Horror Film." *Film Studies* 8: 136-150.

Deleuze, Gilles. 1989. *Cinema 2: The Time-Image*. Translated by Hugh Tomlinson and Robert Galeta. London: Athlone.

___. *Cinema 1: The Movement-Image*. Translated by Hugh Tomlinson and Barbara Habberjam. London: Athlone.

Di Muzio, Gianluca. 2006. "The Immorality of Horror Films." *International Journal of Applied Philosophy* 20 (2): 277-294.

Eldridge, Richard. 2009. "Philosophy In/On/As/And Film: Thomas Wartenberg's *Thinking on Screen: Film as Philosophy*." *Projections* 3 (1): 109-116.

Falzon, Christopher. 2007. *Philosophy Goes to the Movies: An Introduction to Philosophy*. London: Routledge.

Frampton, Daniel. 2006. *Filmosophy*. New York: Columbia University Press.

Freeland, Cynthia A. 2000. *The Naked and the Undead: Evil and the Appeal of Horror*. Boulder: Westview Press.

Gidal, Peter. 1978. "Theory and Definition of Structural/Materialist Film." In *Structural Film Anthology*, edited by Peter Gidal. London: British Film Institute.

Granger, Herbert. 2004. "Cinematic Philosophy in *Le Feu Follet*: The Search for a Meaningful Life." *Film and Philosophy* 8: 74-90.

Grau, Christopher, ed. 2005. *Philosophers Explore The Matrix*. Oxford: Oxford University Press.

Jackson, Kier-La. 2013. *Technology, Monstrosity, and Reproduction in Twenty-First Century Horror*. New York: Palgrave Macmillan.

Jancovich, Mark. 2000. "A Real Shocker': Authenticity, Genre and the Struggle for Distinction." *Continuum* 14 (1): 23-35.

___. 2002. *Horror, The Film Reader*. London: Routledge.

Jarvie, Ian. 1987. *Philosophy of the Film: Epistemology, Ontology, Aesthetics*. New York: Routledge & Kegan Paul.

Knight, Deborah. 2011. "*The Third Man*: Ethics, Aesthetics, Irony." In *Ethics at the Cinema*, edited by Ward E. Jones and Samantha Vice, 285-299. Oxford: Oxford University Press.

Kracauer, Siegfried. 1997. *Theory of Film: The Redemption of Physical Reality*. Princeton: Princeton University Press. First published 1960.

Livingston, Paisley. 2019. "The Bold Thesis Retried: On Cinema as Philosophy." In *Philosophy and Film: Bridging Divides*, edited by Christina Rawls, Diana Neiva, and Steven S. Gouveia, 81-91. New York: Routledge.

___. *Cinema, Philosophy, Bergman: On Film as Philosophy*. Oxford: Oxford University Press.

___. "Recent Work on Cinema as Philosophy." *Philosophy Compass* 3 (4): 590-603.

___. "Theses on Cinema as Philosophy." *Journal of Aesthetics and Art Criticism* 64 (1): 11-18.

Loht, Shawn. 2015. "Phenomenological Preconditions of the Concept of Film-as-Philosophy." *Journal of Aesthetics and Phenomenology* 2 (2): 171-185.

Martin-Jones, David. 2016. "Introduction: Film-Philosophy and a World of Cinemas." *Film-Philosophy* 20 (1): 1-5.

McClelland, Tom. 2019. "Film as Philosophical Thought Experiment: Some Challenges and Opportunities." In *Philosophy and Film: Bridging Divides*, edited by Christina Rawls, Diana Neiva, and Steven S. Gouveia, 92-115. New York: Routledge.

___. "The Philosophy of Film and Film as Philosophy." *Cinema: Journal of Philosophy and the Moving Image* 2: 11-35.

Moura, Vítor. 2006. "The Cinematic Muthos." In *Análises: Actas do 2º Encontro de Filosofia Analítica*, 322-334. Porto: Faculdade de Letras da Universidade do Porto.

___. 2011. "Luzes, câmara ou acção? Três ontologias para a imagem em movimento." *Sem título* 25 (2): 249-275. http://hdl.handle.net/1822/51716.

___. 2019. "Show and Tell: The Identification of Documentary Film." In *The Palgrave Handbook of the Philosophy of Film and Motion Pictures*, edited by Noël Carroll, Laura T. Di Summa, and Shawn Loht, 603–626. Cham: Palgrave Macmillan.

___. 2021. "Art or Intimacy: Assessing a Disjunction." In *Aesthetics, Art and Intimacy*, edited by Carlos João Correia and Emília Ferreira, 3–21. Lisboa: CFUL – Centro de Filosofia da Universidade de Lisboa.

Mulhall, Stephen. 2008 [2002]. *On Film*. 2nd ed. London: Routledge.

Mullarkey, John. 2009. *Philosophy and the Moving Image: Refractions of Reality*. New York: Palgrave Macmillan.

Mulvey, Laura. 1975. "Visual Pleasure and Narrative Cinema." *Screen* 16 (3): 6-18.

Neale, Steve. 2012. "Questions of Genre." In *Film Theory and Criticism*, edited by Leo Braudy and Marshall Cohen, 179-195. Oxford: Oxford University Press.

Rawls, Christina, Diana Neiva, and Steven S. Gouveia, eds. 2019. *Philosophy and Film: Bridging Divides*. New York: Routledge.

Russell, Bruce. 2008. "Film's Limits: The Sequel." *Film and Philosophy* 12: 1-16.

___. 2000. "The Philosophical Limits of Film." *Film and Philosophy* (Special Issue on Woody Allen): 163-167.

Shamir, Tal S. 2016. *Cinematic Philosophy*. Cham: Palgrave Macmillan.

Sinnerbrink, Robert. 2011a. "Re-enfranchising Film: Towards a Romantic Film-Philosophy." In *New Takes in Film-Philosophy*, edited by Havi Carel and Greg Tuck, 25-47. London: Palgrave Macmillan.

___. 2011b. *New Philosophies of Film: Thinking Images*. London: Continuum.

Sitney, P. Adams. 2002. *Visionary Film: The American Avant-Garde, 1943-2000*. 3rd ed. New York: Oxford University Press.

Smith, Murray. 2019. "Response to Neiva." *American Society for Aesthetics Graduate E-Journal* 11 (1).

___. 2016. "Film, Philosophy, and the Varieties of Artistic Value." In *Current Controversies in Philosophy of Film*, edited by Katherine Thomson-Jones, 182-199. London: Routledge.

___. 2006. "Film Art, Argument, and Ambiguity." *Journal of Aesthetics and Art Criticism* 64 (1): 34-42.

Smith, Murray, and Thomas E. Wartenberg. 2006. "Introduction." *Journal of Aesthetics and Art Criticism* 64 (1): 1-9.

Smuts, Aaron. 2009. "Film as Philosophy: In Defense of a Bold Thesis." *Journal of Aesthetics and Art Criticism* 64 (1): 409-420.

Sorfa, David. 2016. "What is Film-Philosophy?" *Film-Philosophy* 20 (1): 1-15.

Stoner, Ian James. 2020. "Moral Psychology and Horror Aesthetics." *The Journal of Aesthetics and Art Criticism* 78 (1): 1-12.

Tietchen, Todd F. 1998. "Samplers and Copycats: The Cultural Implications of the Postmodern Slasher in Contemporary American Film." *Journal of Popular Film and Television* 26 (3): 98-107.

Tudor, Andrew. 2012. "Genre." In Film *Theory and Criticism*, edited by Leo Braudy and Marshall Cohen, 3-11. Oxford: Oxford University Press.

Viegas, Susana. 2018. "Deleuze and Film's Philosophical Value." *Kriterion* 59 (139): 271-286.

___. 2012. "Filosofia do Cinema: Processos de Criação de uma Nova Imagem do Pensamento." PhD diss., Universidade Nova de Lisboa.

Wartenberg, Thomas E. 2017. “The Death of Love: *The Man Who Shot Liberty Valance*’s Dramatization of Philosophy.” Paper presented at the International Conference on Moving Image and Philosophy, Faculdade de Letras da Universidade do Porto.

___. 2016. “Film as Philosophy: The Pro Position.” In *Current Controversies in Philosophy of Film*, edited by Katherine Thomson-Jones, 165-180. London: Routledge.

___. 2011. “On the Possibility of Cinematic Philosophy.” In *New Takes in Film-Philosophy*, edited by Havi Carel and Greg Tuck, 9-24. London: Palgrave Macmillan.

___. 2009. “Response to My Critics.” *Projections* 3 (1): 117-125.

___. 2008. “What Else Films Can Do: A Response to Bruce Russell.” *Film and Philosophy* 12: 27-34.

___. 2007. *Thinking on Screen: Film as Philosophy*. London: Routledge.

___. 2006a. “Beyond *Mere* Illustration: How Films Can Be Philosophy.” *Journal of Aesthetics and Art Criticism* 64 (1): 19-32.

___. 2006b. “Film as Argument.” *Film Studies* 8: 126-137.

___. 2003. “Philosophy Screened: Experiencing *The Matrix*.” *Midwest Studies in Philosophy* 27: 139-152.

Williams, Dan. 2015. *Klein, Sartre and Imagination in the Films of Ingmar Bergman*. Hampshire: Palgrave Macmillan.

Woodcock, Scott. 2013. “The Social Dimensions of Horror.” In *A Companion to the Horror Film*, edited by Harry M. Benshoff, 309-328. Oxford: Wiley-Blackwell.

Thought Experiments

Bishop, Michael A. 1999. “Why Thought Experiments Are Not Arguments.” *Philosophy of Science* 66 (4): 534-541.

Brendel, Elke. 2018. “The Argument View: Are Thought Experiments Mere Picturesque Arguments?” In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach Fehige, and James Robert Brown, 281-293. London: Routledge.

___. 2004. “Intuition Pumps and the Proper Use of Thought Experiments.” *Dialectica* 58: 88-108.

Brown, James R. 2004. “Why Thought Experiments Transcend Empiricism.” In *Contemporary Debates in Philosophy of Science*, edited by Christopher Hitchcock, 23-43. Malden: Blackwell.

___. 2011 [1991]. *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London: Routledge.

Brown, James R. and Yiftach Fehige 2023. “Thought Experiments.” Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy*. Accessed May 10, 2026. https://plato.stanford.edu/entries/thought-experiment/

Buzzoni, Marco. 2018. “Duhem and Mach on Thought Experiments.” *HOPOS: The Journal of the International Society for the History of Philosophy of Science* 8 (1): 1-27.

Clement, J. J. 2009. “The Role of Imagistic Simulation in Scientific Thought Experiments.” *Topics in Cognitive Science* 1 (4): 686-710.

Dancy, Jonathan. 1985. “The Role of Imaginary Cases in Ethics.” *Pacific Philosophical Quarterly* 66 (1-2): 141-153.

Davies, David. 2018. “Art and Thought Experiments.” In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach Fehige, and James Robert Brown, 512-525. New York: Routledge.

Dennett, Daniel C. 2013. *Intuition Pumps and Other Tools for Thinking*. New York: W.W. Norton & Company.

___. 1995. “Intuition Pumps.” In *The Third Culture: Beyond the Scientific Revolution*, edited by John Brockman, 181-197. New York: Simon & Schuster.

Duhem, Pierre. 1913. *La théorie physique: son objet et sa structure*. Paris: Chevalier & Rivière.

Fehige, Yiftach, and Michael T. Stuart. 2014. “On the Origins of the Philosophy of Thought Experiments: The Forerun.” *Perspectives on Science* 22 (2): 179-220.

Frankfurt, Harry G. 1969. “Alternate Possibilities and Moral Responsibility.” *The Journal of Philosophy* 66 (23): 829-839.

Gendler, Tamar S. 2004. “Thought Experiments Rethought—and Reperceived.” *Philosophy of Science* 71 (5): 1152-1163.

___. 2000. *Thought Experiment: On the Power and Limits of Imaginary Cases*. New York: Routledge.

___. 1998. “Galileo and the Indispensability of Scientific Thought Experiments.” *The British Journal for the Philosophy of Science* 49 (3): 397-424.

Gettier, Edmund L. 1963. “Is Justified True Belief Knowledge?” *Analysis* 23 (6): 121-123.

Gooding, D. C. 1992. “What is Experimental about Thought Experiments?” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992 (2): 280-290.

Häggqvist, Sören. 2009. “A Model for Thought Experiments.” *Canadian Journal of Philosophy* 39 (1): 55-76.

Ichikawa, Jonathan, and Benjamin Jarvis, 2009, “Thought-experiment Intuition and truth in Fiction.” *Philosophical Studies* 142: 221–246.

Horowitz, Tamara, and Gerald J. Massey, eds. 1991. *Thought Experiments in Science and Philosophy*. Lanham: Rowman & Littlefield.

Irvine, Andrew D. 1991. “Thought experiments in Scientific Reasoning” in Horowitz and Massey *Thought experiments in Science and* Philosophy. Savage, MD: Rowman & Littlefield Publishers. 149-66.

Jackson, Frank. 1982. “Epiphenomenal Qualia.” *Philosophical Quarterly* 32: 127-136.

Johnson-Laird, Philip N. 1980. “Mental Models in Cognitive Science.” *Cognitive Science* 4 (1): 71-115.

Klee, Robert. 2008. “Physical Scale Effects and Philosophical Thought Experiments.” *Metaphilosophy* 39 (1): 89-104.

Kuhn, Thomas S. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.

Mach, Ernst. 1976 [1905]. *Knowledge and Error: Sketches on the Psychology of Enquiry*. Dordrecht: D. Reidel. Translated by J. McCormack.

Machery, Edouard. 2011. “Thought Experiments and Philosophical Knowledge.” *Metaphilosophy* 42 (3): 191-214.

Miščević, Nadan. 1992. “Mental Models and Thought Experiments.” *International Studies in the Philosophy of Science* 6: 215-226.

Nersessian, Nancy J. 2018. “Cognitive Science, Mental Modeling, and Thought Experiments.” In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach Fehige, and James Robert Brown, 309-326. New York: Routledge.

___. 2013. “Mental Modeling in Conceptual Change.” In *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou, 377-398. Routledge.

___. 2008. *Creating Scientific Concepts*. MIT Press.

___. 2002. “The Cognitive Basis of Model-based Reasoning in Science.” In *The Cognitive Basis of Science*, edited by P. Carruthers, S. Stich, and M. Siegal, 133-153. Cambridge University Press.

___. 1999. “Model-based Reasoning in Conceptual Change.” In *Model-Based Reasoning in Scientific Discovery*, edited by L. Magnani, N. J. Nersessian, and P. Thagard, 5-22. Kluwer/Plenum.

___. 1992a. “How do Scientists Think? Capturing the Dynamics of Conceptual Change in Science.” *Cognitive Models of Science* 15: 3-44.

___. 1992b. “In the Theoretician’s Laboratory: Thought Experimenting as Mental Modeling.” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992 (2): 291-301.

Norton, John D. 2004a. “Why Thought Experiments Do Not Transcend Empiricism.” In *Contemporary Debates in Philosophy of Science*, edited by Christopher Hitchcock, 44-66. Malden: Blackwell.

___. 2004b. “On Thought Experiments: Is There More to the Argument?” *Philosophy of Science* 71, no. 5: 1139–51. https://doi.org/10.1086/425238.

___. 1996. “Are Thought Experiments Just What You Thought?” *Canadian Journal of Philosophy* 26 (3): 333-366.

___. 1991. “Thought Experiments in Einstein’s Work.” In *Thought Experiments in Science and Philosophy*, edited by Tamara Horowitz and Gerald J. Massey, 129-148. Lanham: Rowman & Littlefield.

Peijnenburg, Jeanne, and David Atkinson. 2003. “When are Thought experiments Poor Ones?” *Journal for General Philosophy of Science* 34: 305–322.

Searle, John R. 1980. “Minds, Brains, and Programs.” *Behavioral and Brain Sciences* 3 (3): 417-457.

Sorenson, Roy A. 1992. *Thought Experiments*. Oxford: Oxford University Press.

Stich, Stephen, and Kevin P. Tobia. 2018. “Experimental Philosophy and the Philosophical Tradition.” In *A Companion to Experimental Philosophy*, edited by Justin Sytsma and Wesley Buckwalter, 5-21. Oxford: Wiley-Blackwell.

Stuart, Michael T., Yiftach Fehige, and James Robert Brown, eds. 2018. *The Routledge Companion to Thought Experiments*. New York: Routledge.

Stuart, Michael T. 2016. “Norton and the Logic of Thought Experiments.” *Axiomathes* 26 (4): 451-466.

Thagard, Paul. 2014. “Thought Experiments Considered Harmful.” *Perspectives on Science* 22 (2): 288-305.

Thomson, Judith J. 1971. “A Defense of Abortion.” *Philosophy and Public Affairs* 1: 44-66.

Vidmar, Iris. 2012. “Thought Experiments, Hypotheses, and Cognitive Dimension of Literary Fiction.” *Synthesis Philosophica* 27 (1): 177-193.

Wilkes, Kathleen. 1988. *Real People: Personal Identity without Thought Experiments*. Oxford: Clarendon Press.

Williams, Bernard. 1957. “Personal Identity and Individuation.” *Proceedings of the Aristotelian Society* 57: 229-252.

Williamson, Timothy. 2004. “Philosophical ‘Intuitions’ and Scepticism about Judgement.” *Dialectica* 58 (1): 109-153.

___. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.

___. 2011. “Philosophical Expertise and the Burden of Proof.” *Metaphilosophy* 42 (3): 215-229.

___. Williamson, Timothy. 2016. “Abductive Philosophy.” *Philosophical Forum* 47 (3–4): 263–280.

Wilson, James. 2016. “Internal and External Validity in Thought Experiments.” *Proceedings of the Aristotelian Society* 116 (2): 127-152.

Scientific thought experiments

Colciaghi, Paolo, Yifan Li, Philipp Treutlein, and Tilman Zibold. 2023. “Experiment Shows Einstein-Podolsky-Rosen Paradox Scales Up.” *Phys.org*, June 12, 2023. Accessed April 22, 2025. https://phys.org/news/2023-06-einstein-podolsky-rosen-paradox-scales.html

Einstein, Albert. 1905. “On the Electrodynamics of Moving Bodies.” *Annalen der Physik* 17 (10): 891–921.

Einstein, Albert, Boris Podolsky, and Nathan Rosen. 1935. “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” *Physical Review* 47 (10): 777–780.

Galileo Galilei. 1638. *Discourses and Mathematical Demonstrations Relating to Two New Sciences*. Translated by Henry Crew and Alfonso de Salvio. New York: Macmillan, 1914.

Greene, Jim. 2024. "EPR Paradox." *EBSCO Research Starters*. Accessed April 22, 2025.

Harvard Natural Sciences Lecture Demonstrations. 2025. "Relativity Train." *Harvard University*. Accessed April 22, 2025.
https://sciencedemonstrations.fas.harvard.edu/presentations/relativity-train.

Heisenberg, Werner. 1927. "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik." *Zeitschrift für Physik* 43 (3): 172–198.

Information Philosopher. 2025. "Heisenberg's Microscope." *Information Philosopher*. Accessed April 22, 2025.
MacTutor History of Mathematics Archive. n.d. "Newton's Bucket." University of St Andrews. Accessed April 22, 2025.
https://mathshistory.st-andrews.ac.uk/HistTopics/Newton_bucket/
https://www.informationphilosopher.com/solutions/experiments/heisenberg_microscope/.

Maxwell, James Clerk. 1871. *Theory of Heat*. London: Longmans, Green, and Co.
Norton, John D. n.d. *The Simplest Exorcism of Maxwell's Demon*. University of Pittsburgh. Accessed April 22, 2025.
https://sites.pitt.edu/~jdnorton/Goodies/exorcism_phase_vol/exorcism_phase_vol.html.

Newton, Isaac. 1687. *Philosophiæ Naturalis Principia Mathematica*. London: Royal Society.
Schrödinger, Erwin. 1935. "Schrödinger's Cat." *University of Toronto Physics Department*. Accessed April 22, 2025.
https://faraday.physics.utoronto.ca/PVB/Harrison/SchrodCat/SchrodCat.html.

Schrödinger, Erwin. 1935. "Die gegenwärtige Situation in der Quantenmechanik." *Naturwissenschaften* 23 (49).

Aesthetics and Cognitive Value of Art

Barthes, Roland. 1967. "The Death of the Author." *Aspen* 5/6.

Beardsley, Monroe. 1958. *Aesthetics: Problems in the Philosophy of Criticism*. New York: Harcourt, Brace & World.

Beardsley, Monroe, and William C. Wimsatt. 1946. "The Intentional Fallacy." *The Sewanee Review* 54 (3): 468-88.

Brooks, Cleanth. 1968. "The Heresy of Paraphrase." In *The Well Wrought Urn*. London: Methuen.

Camp, Elisabeth. 2009. "Two Varieties of Literary Imagination: Metaphor, Fiction, and Thought Experiments." *Midwest Studies in Philosophy* 33 (1): 107-130.

Carroll, Noël. 2000. "Interpretation and Intention: The Debate Between Hypothetical and Actual Intentionalism." *Metaphilosophy* 31, no. 1/2: 75–95.

___. 2002. “The Wheel of Virtue: Art, Literature, and Moral Knowledge.” *The Journal of Aesthetics and Art Criticism* 60 (1): 3-26.

Critchley, Simon. 2002. *On Humour*. New York: Routledge.

Cunliffe, Zoë. 2019. “Narrative Fiction and Epistemic Injustice.” *Journal of Aesthetics and Art Criticism* 77 (2): 169–180.

Currie, Gregory. 1990. *The Nature of Fiction*. Cambridge: Cambridge University Press.

___. 2005. *Arts and Minds*. Oxford: Oxford University Press.

Danto, Arthur C. 2005. *The Philosophical Disenfranchisement of Art*. New York: Columbia University Press.

___. 1999. “The Philosopher as Andy Warhol.” In *Philosophizing Art: Selected Essays*, 61-83. Berkeley: University of California Press.

Diamond, Cora. 1991. *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*. Cambridge: MIT Press.

Elgin, Catherine Z. 2019. “Imaginative Investigations: Thought Experiments in Science, Philosophy and Literature”. In *Literature as Thought Experiment* edited by Faulk Bornmuller, Johannes Frazen and Mathis Lessau, 1-16 Paderborn: Wilhelm Fink Verlag (2019),

___. 2014. “Fiction as Thought Experiment.” *Perspectives on Science* 22 (2): 221-241.

___. 2007. “Understanding and the Facts.” *Philosophical Studies* 132 (1): 33-42.

___. 2004. “True Enough.” *Philosophical Issues* 14 (1): 113-131.

___. 2002. “Art in the Advancement of Understanding.” *American Philosophical Quarterly* 39 (1): 1-12.

___. 1993. “Understanding: Art and Science.” *Synthese* 95 (1): 13-28.

Gaut, Berys. 1999. “Identification and Emotion in Narrative Film.” In *Passionate Views: Film, Cognition, and Emotion*, edited by Carl Plantinga and Greg M. Smith, 200-216. Baltimore: Johns Hopkins University Press.

Goodman, Nelson. 1976. *Languages of Art*. Indianapolis: Hackett.

Greenberg, Clement. 1965. “Modernist Painting.” *Art and Literature* 4: 193-201.

Hegel, G. W. F. 1975. *Aesthetics: Lectures on Fine Art*. Vol. 2. Translated by T. M. Knox. Oxford: Clarendon Press.

John, Eileen. 1998. “Reading Fiction and Conceptual Knowledge: Philosophical Thought in Literary Context.” *The Journal of Aesthetics and Art Criticism* 56 (4): 331-348.

Kivy, Peter. 1997. “On the Banality of Literary Truths.” *Philosophical Exchange* 28 (1): 1-11.

Lamarque, Peter, and Stein Haugom Olsen. 1994. *Truth, Fiction, and Literature: A Philosophical Perspective*. Oxford: Clarendon Press.

Levinson, Jerrold. 2010. “Defending hypothetical intentionalism.” *British Journal of Aesthetics* 502: 139-150.

___. 2006. *Contemplating Art: Essays in Aesthetics*. Oxford: Clarendon Press.

Moran, Richard. 1994. "The Expression of Feeling in Imagination." *The Philosophical Review* 103, no. 1: 75–106.

Moura, Vítor. 2019. "Art or Intimacy: Assessing a Disjunction." In *Aesthetics, Art and Intimacy*, edited by Carlos João Correia and Emília Ferreira, 3-21. Lisboa: Centro de Filosofia da Universidade de Lisboa.

Nussbaum, Martha C. 1990. *Love's Knowledge: Essays on Philosophy and Literature*. New York: Oxford University Press.

Stolnitz, Jerome. 1992. "On the Cognitive Triviality of Art." *British Journal of Aesthetics* 12: 191-200.

Walton, Kendall. 1990. *Mimesis as Make-Belief*. Cambridge: Harvard University Press.

General Philosophy

Aristotle. *Rhetoric*. Translated by J.H. Freese. Perseus Digital Library, Tufts University. Accessed July 20, 2025.

https://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0060%3Abook%3D2
Bacon, Francis. 1625. "Of Revenge." Accessed July 21, 2025.
https://www.authorama.com/essays-of-francis-bacon-5.html.

Baudrillard, Jean. 1998. *Simulacra and Simulation*. Translated by Sheila Faria Glaser. Ann Arbor: University of Michigan Press.

Blackburn, Simon. 2004. "Foreword." In *What Philosophy Is*, edited by Havi Carel and David Gamez. London: Continuum.

Butler, Rex. 2002. *Jean Baudrillard: The Defence of the Real*. London: Sage Publications.

Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Davidson, Donald. 1987. "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association* 60 (3): 441-458.

Dennett, Daniel. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: MIT Press.

___. 1978. "Where Am I?" In *Brainstorms: Philosophical Essays on Mind and Psychology*, 310–323. Montgomery, VT: Bradford Books.

Descartes, René. 1996. *Meditations on First Philosophy*. Translated by John Cottingham. Cambridge: Cambridge University Press.

Dietrich, Eric. 2011. "There Is No Progress in Philosophy." *Essays in Philosophy* 12 (2): Article 9.

Fletcher, George P. 2000. *Rethinking Criminal Law*. Oxford: Oxford University Press.

Frankfurt, Harry G. 1969. “Alternate Possibilities and Moral Responsibility.” *The Journal of Philosophy* 66 (23): 829–839.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. New York: Oxford University Press.

Gutting, Gary. 2009. *What Philosophers Know: Case Studies in Recent Analytic Philosophy*. New York: Cambridge University Press.

Jackson, Frank. 1982. “Epiphenomenal Qualia.” *The Philosophical Quarterly* 32 (127): 127–136.

Jung, C. G. 1969. *Archetypes and the Collective Unconscious*. Vol. 9, pt. 1 of *The Collected Works of C. G. Jung*, edited by Herbert Read, Michael Fordham, and Gerhard Adler; translated by R. F. C. Hull. Princeton, NJ: Princeton University Press. Accessed July 26, 2025. https://www.jungiananalysts.org.uk/wp-content/uploads/2018/07/C.-G.-Jung-Collected-Works-Volume-9i_-The-Archetypes-of-the-Collective-Unconscious.pdf.

Laruelle, François. 1999. “A Summary of Non-Philosophy.” *Pli* 8: 138-148.

Locke, John. 1690. *An Essay Concerning Human Understanding*. Edited by Peter H. Nidditch. Oxford: Clarendon Press, 1975.

Marx, Karl. 1988. *Economic and Philosophic Manuscripts of 1844*. Translated by Martin Milligan. Amherst, NY: Prometheus Books. First published 1844.

Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Overgaard, Søren, Paul Gilbert, and Stephen Burwood. 2013. *An Introduction to Metaphilosophy*. Cambridge: Cambridge University Press.

Parfit, Derek. 2016. “Divided Minds and the Nature of Persons.” In *The Blackwell Companion to Consciousness*, edited by Susan Schneider, ch. 8. Hoboken, NJ: Wiley-Blackwell.

Persson, Ingmar. 2018. “The Fundamental Problem of Philosophy: Its Point.” *Journal of Practical Ethics* 6 (1): 52–68.

Platão. 2001. *A República*. 9th ed. Translated by Maria Helena da Rocha Pereira. Lisboa: Fundação Calouste Gulbenkian.

Plato. *Republic*. Translated by Benjamin Jowett. *Perseus Digital Library*, Tufts University. Accessed July 20, 2025. https://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0168%3Abook%3D7%3Asection%3D515c

Putnam, Hilary. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.

___. 1975. “The Meaning of ‘Meaning’.” In *Mind, Language and Reality: Philosophical Papers, Volume 2*, 215–271. Cambridge: Cambridge University Press.

___. 1964. “Robots: Machines or Artificially Created Life?” *The Journal of Philosophy* 61 (21): 668–691.

Quine, W. V. O. 1960. *Word and Object*. Cambridge: MIT Press.

Rawls, John. 1971. *A Theory of Justice*. Cambridge: Belknap Press.

Rescher, Nicholas. 1985. *The Strife of Systems*. Pittsburgh: University of Pittsburgh Press.

___. 2006. *Collected Papers, Volume IX: North and South America*. Rutgers University Studies in Metaphilosophy. New Brunswick, NJ: Transaction Books.

Rooney, Phyllis. 2022. "Feminist Perspectives on Argumentation." Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy*. Accessed July 29, 2025. https://plato.stanford.edu/entries/feminism-argumentation/.

Searle, John R. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 417–457.

Scheiter, Krisanna M. 2010. "Aristotle on the Purpose of Revenge". In *Best Served Cold: Studies on Revenge*, edited by Sheila C. Bibb and Daniel Escandell, 1-12. Leiden: Brill.

Turing, Alan M. 1950. "Computing Machinery and Intelligence." *Mind* 59 (236): 433–460.

Walen, Alec. "Retributive Justice." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University, 2021. Accessed July 20, 2025. https://plato.stanford.edu/entries/justice-retributive/.

Williams, Bernard. 1957. "Personal Identity and Individuation." *Proceedings of the Aristotelian Society* 57: 229–252.

Literature on Wes Craven and his films

Berns, Fernando Gabriel Pagnoni and John Darowski. 2023. *A Critical Companion to Wes Craven*. Lanham: Lexington Books.

Brehmer, Noah. 2021. "The Reality of Storytelling in *Wes Craven's New Nightmare*." *Medium*. Accessed July 26, 2025. https://nbrehmer.medium.com/the-reality-of-storytelling-in-wes-cravens-new-nightmare-bca7b8941264

Bukta, Ivan. 2018. "Wes Craven: Whatever You Do, Don't Fall Asleep." *Horror Obsessive*. Accessed July 23, 2025. https://horrorobsessive.com/2018/08/31/wes-craven-whatever-you-do-dont-fall-asleep/.

Callahan, Corey. 2020. "Moral Equivalence in Wes Craven's *The Last House on the Left*, *The Hills Have Eyes*, and *A Nightmare on Elm Street*." *Horror Obsessive*. Accessed July 2021, 2025. https://horrorobsessive.com/2020/10/09/moral-equivalence-in-wes-cravens-the-last-house-on-the-left-the-hills-have-eyes-and-a-nightmare-on-elm-street/.

Catt, Chris. 2002. "Wes Craven Movies, Ranked/Analyzed." *Creepy Catalog*. Accessed July 20, 2025. https://creepycatalog.com/wes-craven-movies/.

Chaka, Chris. 2020. "Only a Movie: The Lasting Impact of The Last House on the Left." *VHS Revival*, Accessed July 21, 2025. https://vhsrevival.com/2020/10/02/only-a-movie-the-lasting-impact-of-the-last-house-on-the-left/.

Craven, Wes. 1982 [2019]. Interview by Alan Jones. *In Wes Craven: Interviews*, edited by Shannon Blake Skelton. Jackson: University Press of Mississippi.

___. 1984. Interview by Glenn Lovell. “Wes Craven: Professor of Horror”. *CinemaDope*. Accessed July 20, 2025. https://cinemadope.com/the-professor-of-horror/.

___. 1988 [2019]. Interview by Dennis Fischer. In *Wes Craven: Interviews*, edited by Shannon Blake Skelton. Jackson: University Press of Mississippi.

___. 1990 [2019] Interview by Michael Banka. In *Wes Craven: Interviews*, edited by Shannon Blake Skelton. Jackson: University Press of Mississippi.

___. 1997. Interview by Charlie Rose. *Charlie Rose*, February 17, 1997. Video. Accessed July 24, 2025. https://charlierose.com/videos/8535.

___. 1999 [2019] Interview by Robert J. Emery. In *Wes Craven: Interviews*, edited by Shannon Blake Skelton. Jackson: University Press of Mississippi.

___. 2005 [2019]. Interview by Tony Timpone. In *Wes Craven: Interviews*, edited by Shannon Blake Skelton. Jackson: University Press of Mississippi.

___. 2015. [2019] Interview by Jennifer Juniper Stratford. In *Wes Craven: Interviews*, edited by Shannon Blake Skelton. Jackson: University Press of Mississippi.

Flasz, Jakub. 2020. “The Wes Craven Retrospective: *The Last House On The Left* (1972).” *Flasz on Film.* Accessed July 21, 2025. https://flaszonfilm.com/2020/11/04/the-wes-craven-retrospective-the-last-house-on-the-left-1972/.

Housman, Andrew. 2020. “Last House on the Left: Wes Craven’s Director Roots Explained.” *Screen Rant*. Accessed July 21, 2025. https://screenrant.com/last-house-left-movie-wes-craven-director-roots-explained/.

Johnson, Rachel and Patrick Hayes. 2003. “Why Wes Craven is One of the Greatest Horror Movie Directors”. *MovieWeb*. Accessed July 20, 2025. https://movieweb.com/greatest-horror-movie-director.

Keiper, Brian. 2023. “Rubber Reality and Its Variations in the Films of Wes Craven.” In *A Critical Companion to Wes Craven*, edited by Fernando Gabriel Pagnoni Berns and John Darowski, chapter 16. Lanham: Lexington Books.

Mitchell, Sean. 1990. “The X Rating Gets Its Day in Court: Movies: William Kunstler will argue today that the X for ‘Tie Me Up!’ violates New York law. The MPAA counters that its system thwarts government censorship.” *Los Angeles Times*, June 21, 1990. Accessed July 26, 2025. https://www.latimes.com/archives/la-xpm-1990-06-21-ca-535-story.html

Muir, John K. 2002. *Horror Films of the 1970s*. New Jefferson, NC: McFarland.

___. 2004. *Wes Craven: The Art of Horror*. Jefferson, NC: McFarland.

Nunez, John Paul. 2022. “How Wes Craven’s *New Nightmare* Subtly Criticizes the Previous Sequels.” *Horror Obsessive*, December 29, 2022. Accessed July 26, 2025. https://horrorobsessive.com/2022/12/29/how-wes-cravens-new-nightmare-subtly-criticizes-the-previous-sequels/

Shapiro, Marc. 1994. “Wes Craven’s Psycho Analysis.” *Fangoria*, no. 138: 28–31.

Skelton, Shannon Blake. 2019. *Wes Craven: Interviews*. Jackson: University Press of Mississippi.

Filmography

A Nightmare on Elm Street. Directed by Wes Craven. 1984. Los Angeles, CA: New Line Cinema.

A Simple Plan. Directed by Sam Raimi. 1998. Los Angeles, CA: Paramount Pictures.

Alien. 1979. Directed by Ridley Scott. Los Angeles: 20th Century Fox.

Aliens. 1986. Directed by James Cameron. Los Angeles: 20th Century Fox.

Alien 3. 1992. Directed by David Fincher. Los Angeles: 20th Century Fox.

Alien Resurrection. 1997. Directed by Jean-Pierre Jeunet. Los Angeles: 20th Century Fox.

Alien: Covenant. 2017. Directed by Ridley Scott. Los Angeles: 20th Century Fox.

All of Me. Directed by Carl Reiner. 1984. Los Angeles, CA: Universal Pictures.

Baby Doll. Directed by Elia Kazan. 1956. Burbank, CA: Warner Bros.

Blade Runner. Directed by Ridley Scott. 1982. Burbank, CA: Warner Bros.

Crimes and Misdemeanors. Directed by Woody Allen. 1989. Los Angeles, CA: Orion Pictures.

Empire. Directed by Andy Warhol. 1965. New York: Andy Warhol Films.

Eternal Sunshine of the Spotless Mind. Directed by Michel Gondry. 2004. Los Angeles, CA: Focus Features.

Friday the 13th. 1980. Directed by Sean S. Cunningham. Hollywood: Paramount Pictures.

Halloween. 1978. Directed by John Carpenter. Los Angeles: Compass International Pictures.

Hiroshima mon amour. Directed by Alain Resnais. 1959. Paris: Pathé Films.

Interstellar. Directed by Christopher Nolan. 2014. Burbank, CA: Warner Bros. Pictures.

It Happened One Night. Directed by Frank Capra. 1934. Culver City, CA: Columbia Pictures.

October [*Ten Days That Shook the World*]. Directed by Sergei Eisenstein. 1928. Moscow: Sovkino.

Psycho. 1960. Directed by Alfred Hitchcock. Hollywood: Paramount Pictures.

Pumpkinhead. Directed by Stan Winston. 1988. Los Angeles: MGM.

Scream. 1996. Directed by Wes Craven. Los Angeles: Dimension Films.

Scream 2. 1997. Directed by Wes Craven. Los Angeles: Dimension Films.

Scream 3. 2000. Directed by Wes Craven. Los Angeles: Dimension Films.

Scream *4*. 2011. Directed by Wes Craven. Los Angeles: Dimension Films.

Scream. 2022. Directed by Matt Bettinelli-Olpin and Tyler Gillett. Hollywood: Paramount Pictures.

Scream *VI*. 2023. Directed by Matt Bettinelli-Olpin and Tyler Gillett. Hollywood: Paramount Pictures.

Shocker. Directed by Wes Craven. 1989. Los Angeles, CA: Universal Pictures.

The Act of Killing. Directed by Joshua Oppenheimer. 2012. London: Dogwoof Pictures.

The Cave: An Adaptation of Plato's Allegory in Clay. Directed by Michael Ramsey. 2008.

The Good Place. Created by Michael Schur. 2016-2020. Los Angeles, CA: NBC.

The Hills Have Eyes. Directed by Wes Craven. 1977. Los Angeles, CA: Blood Relations Co.

The Invisible Man. Directed by Leigh Whannell. 2020. Universal Pictures.

The Last House on the Left. Directed by Wes Craven. 1972. New York: Hallmark Releasing Corporation.

The Man Who Shot Liberty Valance. 1962. Directed by John Ford. Hollywood: Paramount Pictures.

The Matrix. Directed by Lana and Lilly Wachowski. 1999. Burbank, CA: Warner Bros.

The New World. Directed by Terrence Malick. 2005. Los Angeles, CA: New Line Cinema.

The People Under the Stairs. Directed by Wes Craven. 1991. Los Angeles, CA: Universal Pictures.

The Pervert's Guide to Ideology. Directed by Sophie Fiennes. 2012. London: P Guide Productions.

The Terminator. 1984. Directed by James Cameron. Los Angeles: Hemdale Film Corporation.

The Texas Chain Saw Massacre. 1974. Directed by Tobe Hooper. Dallas: Vortex Inc.

The Thing. 1982. Directed by John Carpenter. Los Angeles, CA: Universal Pictures.

The Third Man. Directed by Carol Reed. 1949. London: London Films.

The Twilight Zone, "The Little People." Directed by William Claxton. 1962. CBS Television, Season 3, Episode 28.

The Virgin Spring. Directed by Ingmar Bergman. 1960. Stockholm: Svensk Filmindustri.

Three Billboards Outside Ebbing, Missouri. Directed by Martin McDonagh. 2017. Fox Searchlight Pictures.

Vertigo. Directed by Alfred Hitchcock. 1958. Hollywood, CA: Paramount Pictures.

Waking Life. Directed by Richard Linklater. 2001. Los Angeles, CA: Fox Searchlight Pictures.

Ways of Seeing. Directed by Michael Dibb. 1972. London: BBC Television.

Wes Craven's New Nightmare. Directed by Wes Craven. 1994. Los Angeles, CA: New Line Cinema.

With Babies and Banners. Directed by Lorraine Gray. 1977. New Day Films.

Figure sources

Alex on Film. 2014. "The Last House on the Left (1972)." Acessed June 20, 2025. https://alexonfilm.com/2014/06/04/the-last-house-on-the-left-1972/.

Bochenek, Annette. 2025. "Smile: A Reflection on Modern Times." *Hometowns to Hollywood*. Accessed April 26, 2025. https://hometownstohollywood.com/film-reflections/films-of-the-1930s/smile-a-reflection-on-modern-times/.

Empire State Building. 2025. "EXIBIÇÃO EXCLUSIVA: 'Império' de Andy Warhol completa 60 anos!" *Empire State Building*, accessed April 26, 2025. https://www.esbnyc.com/pt/blog/exclusive-screening-andy-warhols-empire-turns-60.

Experimental Cinema. 2025. “Cinemateca Portuguesa: Os Filmes de Ernie Gehr.” Accessed April 26, 2025. https://expcinema.org/site/en/events/cinemateca-portuguesa-os-filmes-de-ernie-gehr.

Irshad, Huzaifa. 2024. “Beyond the Spin: Newton’s Bucket and the Riddle of Spacetime.” *Journal of Young Physicists*. Accessed March 26, 2025. https://www.journalofyoungphysicists.org/post/beyond-the-spin-newton-s-bucket-and-the-riddle-of-spacetime.

New Scientist. 2015. “Ava: Ex Machina: Quest to Create an AI Takes No Prisoners.” *New Scientist*. Accessed April 27, 2025. https://www.newscientist.com/article/mg22530054-600-ex-machina-quest-to-create-an-ai-takes-no-prisoners/.

Norton, John D. n.d. *The Simplest Exorcism of Maxwell’s Demon*. University of Pittsburgh. Accessed April 22, 2025. https://sites.pitt.edu/~jdnorton/Goodies/exorcism_phase_vol/exorcism_phase_vol.html.

Physical Review X. 2023. “Einstein-Podolsky-Rosen Paradox at Scales.” June 2023. Accessed April 22, 2025. https://phys.org/news/2023-06-einstein-podolsky-rosen-paradox-scales.html.

Schrödinger, Erwin. 1935. “Schrödinger’s Cat.” University of Toronto Physics Department. Accessed April 22, 2025. https://faraday.physics.utoronto.ca/PVB/Harrison/SchrodCat/SchrodCat.html.

The Take. n.d. “Was ‘Rear Window’ Shot in an Actual Apartment Courtyard or an Elaborate Set?” *The Take*. Accessed April 27, 2025. https://the-take.com/read/was-arear-windowa-shot-in-an-actual-apartment-courtyard-or-an-elaborate-set.

VHS Revival. 2020. “Only a Movie: The Lasting Impact of The Last House on the Left.” *VHS Revival*. Accessed June 22, 2025. https://vhsrevival.com/2020/10/02/only-a-movie-the-lasting-impact-of-the-last-house-on-the-left/.

Warhol, Andy. 2015. *Poster: The Last House on the Left*. *IMDb*. Accessed June 20, 2025. https://www.imdb.com/title/tt0068833/mediaviewer/rm3095408385/?ref_=tt_ov_i.