Universidade do Minho
Escola de Psicologia

Maria Inês Abreu Fortes

Suboptimal choice and the value of information

março de 2016

Universidade do Minho
Escola de Psicologia

Maria Inês Abreu Fortes

Suboptimal choice and the value of information

Tese de Doutoramento em Psicologia Básica

Trabalho efetuado sob a orientação do
Professor Doutor Armando Machado
e do
Professor Doutor Marco Vasconcelos

março de 2016

DECLARAÇÃO

Nome: Maria Inês Abreu Fortes

Endereço electrónico: ines.fortes@gmail.com

Número do Bilhete de Identidade: 13186505

Título da tese: Suboptimal choice and the value of information

Orientadores:
Professor Doutor Armando Machado
Professor Doutor Marco Vasconcelos

Ano de conclusão: 2016

Designação do Doutoramento: Psicologia Básica

Universidade do Minho, 30 de Março de 2016

Assinatura: Maria Inês Abreu Fortes

STATEMENT OF INTEGRITY

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration. I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, March 30, 2016.

Full name: Maria Inês Abreu Fortes

Signature: Maria Inês Abreu Fortes

Aos 'meus'

ACKNOWLEDGEMENTS

Aos meus orientadores. Obrigada ao Professor Armando Machado que com o seu entusiasmo instigou em mim o gosto por esta área. Obrigada por confiar na minha autonomia, por incentivar-me a ir mais além, e por ensinar os seus alunos a não se satisfazerem com algo que pode ser melhorado. Obrigada Professor Marco Vasconcelos, por me abrir os horizontes e ter apresentado um lado mais “biológico” dentro da psicologia. Mas mais importante, por acreditar nas minhas ideias, pelo apoio e entusiasmo neste projecto.

Thank you to Dr. Thomas Zentall, for welcoming me in your lab. I am deeply grateful for everything you shared, for always being available, and for the long and motivating discussions. Thank you for being a mentor and a friend. My stay in your lab definitely improved me as a scientist. Thank you Aaron for being so welcoming and accepting. For the long talks and discussions, and for making me feel more like at home. I am sure without you my stay would not have been as productive and enjoyable.

Obrigada aos professores e colegas de laboratório, pela ajuda e por todos os comentários e sugestões. Obrigada por todas as vezes em que cuidaram dos meus pombos e das minhas experiências como se fossem vossas. Aos colegas que colaboraram comigo na Uminho, Luis Cabello, Miguel Herves, Sara Teixeira, Catarina Soares, Margarida Vasconcelos e Álvaro Viudez; aprendi muito com cada um de vós. Aos colegas e amigos, por acreditarem em mim, por vezes mais que eu própria. À Eugénia, Andréia, Thaís, Catarina, Álvaro, Nocas, Guida e Tatiana, pelo apoio, por me fazerem seguir em frente. Talvez não saibam, mas a vossa energia deu-me muitas vezes força.

Ao Carlos, por absolutamente tudo. Por tanto me ouvires, por tanto me ensinares. Pelas longas discussões que tornaram esta tese o que é hoje. Mas acima de tudo por nunca me deixares caminhar só, e por tornares a caminhada mais leve. Obrigada por me obrigares a parar, respirar fundo e fazer ver que há coisas mais importantes na vida.

A toda a minha família! Por me apoiarem sempre, sempre. Por toda a paciência que têm comigo (que sei que é muita!), por toda a compreensão. Em especial aos meus pais, porque acredito que foi ao ensinarem-me a questionar, que estou onde estou hoje. À minha mãe por estar sempre comigo, por ser para mim um exemplo, como mulher e como profissional. Ao meu pai por tentar sempre ajudar (mesmo quando tal é impossível) e fazer de mim alguém mais forte. À minha mãe e ao João, pelas nossas fugas a três que tantas vezes me fizeram descer à terra.

À Ema por estar sempre ao meu lado. À Zoe pelas alegrias que trouxe.

Obrigada à Fundação para a Ciência e Tecnologia por me ter financiado nestes últimos quatro anos (SFRH/BD/77061/2011). Este apoio foi fundamental.

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

POPH
PROGRAMA OPERACIONAL POTENCIAL HUMANO

QREN
QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
PORTUGAL 2007.2013

Governo da República Portuguesa

UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

SUBOPTIMAL CHOICE AND THE VALUE OF INFORMATION

The environment is full of unpredictable events. Information that reduces uncertainty about events allows an organism to better predict and prepare to what is to come. For that reason, obtaining information can be crucial for survival. In the present dissertation we explored a task in which animals trade food for information. In this task, animals choose between two options: the Informative Option delivers food on 20% of the trials after a 10-s delay, signaled by “good-news stimulus”, and delivers no food on the remaining 80% of the trials, signaled by a “bad-news stimulus”. The Non-informative Option delivers food after 10 s on 50% of the trials, regardless of which of two different stimulus is shown. A consistent and almost exclusive preference for the Informative Option has been found, even though that results in loss of food. In Study 1 we showed that this suboptimal choice can be explained by animals ignoring the bad-news stimulus: its probability and duration had little to no effect on preference. In Study 2 we showed that, when given the opportunity, animals escape from the bad-news stimulus. Moreover, our results suggest that ignoring the bad-news stimulus means that, even though this stimulus is perceived, it is not associated with the choice of the Informative Option. In Study 3 we increased the probability of reinforcement following the until-then bad-news stimulus, and found that, as paradoxically as it may seem, within the tested range, the value of the Informative Option decreased as the probability of reinforcement increased. This increase in reinforcement was consistent with an increase in the association of the bad-news stimulus to the Informative Option. In all studies we showed that an optimal foraging-based model (Vasconcelos, Monteiro, & Kacelnik, 2015) accounted at least as well as a conditioned reinforcement-based model (Mazur, 1987), thus deconstructing the idea that this suboptimal behavior contradicts optimal foraging principles. These findings are expected to contribute in calling attention to the ecology of the animal, bringing animal psychology and behavior ecology closer.

ESCOLHAS SUBÓTIMAS E O VALOR DA INFORMAÇÃO

O meio ambiente é repleto de eventos imprevisíveis. A informação que reduz a incerteza acerca dos eventos permite um organismo prever e preparar-se melhor para o futuro. Por essa razão, obter informação pode ser crucial para a sobrevivência. Nesta dissertação explorámos uma tarefa na qual os animais trocam comida por informação. Nesta tarefa, os animais escolhem entre duas opções: a Opção Informativa dá comida em 20% dos ensaios após um atraso de 10 s, sinalizados por um “estímulo de boas notícias”, e não dá comida nos restantes 80% dos ensaios, sinalizados por um “estímulo de más notícias”. A Opção Não Informativa dá comida após 10 s em 50% dos ensaios, independentemente de qual de dois estímulos é apresentado. Tem-se encontrado uma preferência consistente e praticamente exclusiva pela Opção Informativa, apesar disso resultar em perda de comida. No Estudo 1 mostrámos que esta escolha subótima pode ser explicada pelos animais ignorarem o estímulo das más notícias: a sua probabilidade e duração tiveram pouco ou nenhum efeito na preferência. No Estudo 2 mostrámos que, quando dada a oportunidade, os animais escapam do estímulo das más notícias. Para além disso, os nossos resultados sugerem que ignorar o estímulo das más notícias significa que, apesar de este estímulo ser percecionado, não é associado com a escolha da Opção Informativa. No Estudo 3 aumentámos a probabilidade de reforço após o até então estímulo das más notícias e descobrimos que, paradoxalmente, dentro da amplitude testada, o valor da Opção Informativa diminuiu à medida que a probabilidade de reforço aumentou. Este aumento foi consistente com um aumento na associação do estímulo das más notícias com a Opção Informativa. Em todos os estudos mostrámos que um modelo baseado na em *optimal foraging* (Vasconcelos, Monteiro, & Kacelnik, 2015) dá conta dos dados tanto quanto um modelo baseado em reforçamento condicionado (Mazur, 1987), desconstruindo assim a ideia de que este comportamento subótimo contradiz os princípios de *optimal foraging*. Esperamos que estes resultados contribuam para chamar à atenção a ecologia do animal, unindo assim a psicologia animal e a ecologia comportamental.

TABLE OF CONTENTS

CHAPTER I — INTRODUCTION

CHAPTER II — ON THE STRUCTURE AND ROLE OF OPTIMALITY MODELS IN THE STUDY OF BEHAVIOR

CHAPTER III ─ STUDY 1: IGNORING "BAD NEWS": PIGEONS DISREGARD STIMULI THAT ARE NEVER FOLLOWED BY FOOD

CHAPTER IV ─ STUDY 2: DO PIGEONS ESCAPE FROM "BAD NEWS"?

CHAPTER V ─ STUDY 3: THE PARADOXICAL EFFECT OF LOWER REWARD PROBABILITY

CHAPTER VI – CONCLUSION

APPENDICES

ABBREVIATIONS, ACRONYMS AND SYMBOLS

A	amount
ANOVA	Analysis of Variance
CI	Confidence Interval
D	delay
EoR	Expectation of the Ratios
FI	fixed-interval schedule of reinforcement
G(t)	Gain function
h	height
HDM	Hyperbolic Discounting Model
ITI	Intertrial Interval
K	discounting rate parameter
l	length
LL	Larger later reward
M	Mean
p	probability
MVT	Marginal Value Theorem
R	rate
RoE	Ratio of Expectations
RRM	Reinforcement Rate Model
SCM	Sequential Choice Model
$\mathbf{S_p}$	Stimulus followed by food with probability p
SS	Smaller shorter reward
V	Value
SEM	Standard Error of the Mean
w	width

FIGURES

CHAPTER I — INTRODUCTION

Figure 1. Concurrent-chains procedure used by Kendall (1974).

Figure 2. Schematic of the design used in Gipson et al. (2009), left panel, and in Stagner and Zentall (2010), right panel.

Figure 3. Schematic of the design used in Zentall and Stagner (2011a).

Figure 4. Schematic of the design used in Zentall and Stagner (2011b).

CHAPTER II — ON THE STRUCTURE AND ROLE OF OPTIMALITY MODELS IN THE STUDY OF BEHAVIOR

Figure 1. Graphical representation of the marginal value theorem as applied to a central place forager. The origin of coordinates is set at the arrival time to a patch. The horizontal axis shows the total travel time (round trip) growing to the left and patch time growing to the right; the vertical axis shows accumulated food, with the curvilinear function showing total gains as a function of patch time. An animal leaving all patches after collecting food for a time t will experience an overall rate of returns R, given by the ratio of accumulated gains to the sum of travel and patch times. In (a) several potential leaving times are shown. The line with slope $R(t_{op})$ represents the maximum rate of prey acquisition, and serves to identify the optimal patch leaving time. Also shown for comparison are two alternative strategies, with rates $R(t_a)$ and $R(t_b) = R(t_c)$, both less profitable than $R(t_{op})$; (b) shows the effect of travel time. When the round trip travel decreases so does the optimal patch residence time and consequently the optimal load size. Notice that only at the optimal departure time is the slope of G(t) identical to the overall rate R.

Figure 2. Load size as a function of round trip times in a field experiment with starlings as subjects. Each dot represents the average number of mealworms collected over approximately 50 trips to the same foraging site; the line represents the predicted optimal

number of prey according to the MVT when physiological costs are considered. Adapted from Kacelnik (1984).

Figure 3. The effect of travel time and intertrial-interval (ITI) on decisions. (a) In the depleting patch problem, the decision being modeled is patch departure, which marks the start of a new foraging cycle. Travel time occurs between the decision, indicated by the two black dots, and a new patch arrival. Given the cyclical nature of the problem, the two dots mark the same point in time; (b) in the discounting problem, the decision between two options and the programmed delay occurs between the decision, occurring at the moment indicated by the black dot, and either outcome. Travel time (or ITI in lab simulations) occurs between outcomes and new choices. Although the overall rate of reward, as indicated by the slope of the broken lines, shows that travel time has the same effect on rate of reward in both the depleting patch and the discounting problems, reinforcement analysis expects them to be substantially different, and predicts that travel time will control decisions in the MVT while ITI will be irrelevant in the discounting problem.

Figure 4. Median latencies to accept each option during sequential encounters as a function of the option’s profitability and that of the alternative option that could potentially be encountered in the same environment, averaged across subjects. Adapted from Shapiro et al. (2008).

Figure 5. Illustration of the SCM ability to predict simultaneous choice from sequential encounters. (a) Obtained versus predicted proportion of choices for the option yielding higher rate of reward (always labelled A) according to the SCM. Each rate is represented by a different symbol. Two linear regression lines are included. The dashed line corresponds to an unconstrained regression and the doted one is constrained to pass through the origin (adapted from Shapiro et al., 2008); (b) Average proportion of choices accurately predicted by the SCM (solid circles, left axis, molecular predictions), and difference between observed and predicted preference strength (open circles, right axis, molar predictions) as a function of the number of preceding sequential trials used to predict each choice in simultaneous presentations (adapted from Vasconcelos et al., 2010)

CHAPTER III ─ STUDY 1: IGNORING "BAD NEWS": PIGEONS DISREGARD STIMULI THAT ARE NEVER FOLLOWED BY FOOD

CHAPTER IV ─ STUDY 2: DO PIGEONS ESCAPE FROM “BAD NEWS”?

CHAPTER V ─ STUDY 3: THE PARADOXICAL EFFECT OF LOWER REWARD PROBABILITY

Figure 7. Obtained (black dots) and predicted indifference points by the Hyperbolic Discounting Model (white triangles) and the Reinforcement Rate Model (white dots) as a function of the probability of reinforcement following the S_p stimulus. A value of D = 20 s for the informative terminal-link delay was used in all predictions, and for the Hyperbolic Discounting Model K was set to 1.

TABLES

CHAPTER III – STUDY1: IGNORING "BAD NEWS": PIGEONS DISREGARD STIMULI THAT ARE NEVER FOLLOWED BY FOOD

CHAPTER I

INTRODUCTION

1. The observing studies

1.1. The Wyckoff's procedure

The experiments of L. B. Wyckoff's doctoral dissertation, published in 1969 in an edited book titled "Conditioned Reinforcement", initiated a line of research now known as *observing behavior*. At that time many researchers were studying discrimination learning, that is, how animals learn to discriminate and respond to different stimuli. For instance, how they learn to respond to positive stimuli (i.e., stimuli associated with reward), and not respond to negative stimuli (i.e., stimuli associated with the absence of reward).

In order to learn a discrimination, animals have to observe the stimuli or to "contact them with their sensory receptors" (Browne & Dinsmoor, 1974, p. 165). Wyckoff (1969) wanted to study what was the role of observing the stimuli in learning a discrimination, and defined observing as "a response which results in exposure to a pair of discriminative stimuli" (Wyckoff, 1969, p. 237). Because the more traditional measure of observing, such as looking and orienting towards the stimuli, were difficult to measure, Wyckoff designed a procedure that allowed him to objectively measure observing. In his experiments, pigeons were exposed to a series of 30-s trials in which a white key was present. Even though the key was always illuminated with white, there were two intermixed trial types: positive and negative trials. In positive trials, the first peck after the 30-s period elapsed was reinforced with access to food - a *Fixed Interval* 30 s schedule of reinforcement (FI 30); in negative trials, pecks were never reinforced, (i.e., an extinction schedule was in effect). Concurrently with the white key, pigeons had a pedal in the floor that, while pressed, changed the key color to signal which trial was in effect: red when it was a positive trial and green when it was a negative trial. Thus, pressing the pedal was the observing response and Wyckoff was interested in knowing if and how animals would learn this observing response. Importantly, the programmed probability of reinforcement was independent of observing, that is, the only effect of pressing the pedal was to inform which schedule of reinforcement was in effect.

Wyckoff (1969) found that animals pressed significantly more the pedal when it produced stimuli correlated with the schedule of reinforcement currently in effect, than when it produced stimuli uncorrelated with the current schedule. The author concluded that the discriminative stimuli served as conditioned reinforcers, which reinforced the observing response.

The major question raised by this study was why animals would press a pedal just to obtain a discriminative stimulus given that this response was not associated with an increase in primary reinforcement. One could hypothesize that because food delivery depended on the animal's response (in positive trials at least one response was required), obtaining information about the schedule in effect was advantageous because it allowed the animal not to spend energy pecking a key when food was not available. However, even when food is response-independent, animals press a pedal or peck a key to obtain discriminative stimuli (e.g., Browne & Dinsmoor, 1974).

Because the discriminative stimulus produced by the observing response *informed* the animal about the schedule of reinforcement currently in effect, some authors hypothesized that information itself could be reinforcing. This hypothesis was appealing because, on the one hand, it was reasonable and intuitive and, on the other hand, it was amenable to quantitative analyses mainly derived from *information theory* (Shannon & Weaver, 1949). The latter presupposes that positive and negative stimuli should be equally important in reinforcing the observing response because they are equally informative (i.e., both are perfectly correlated with the presence and absence of food, respectively). However, after Wyckoff's seminal work, several studies (e.g., Browne & Dinsmoor, 1974; Dinsmoor, Browne, Lawrence, & Wasserman, 1971) have shown that the emission of the observing response is selective: when animals observe and obtain a positive stimulus, they continue with the observing response, whereas when they obtain a negative stimulus they stop the observing response (Dinsmoor, Browne, Lawrence, 1972). This result suggests that, contrarily to what is assumed by the information hypothesis, only stimuli associated with food maintain the observing behavior, serving as conditioned reinforcers for this behavior (see Dinsmoor, 1983 for further discussion). After this finding, much research was devoted to unravel the reasons for the asymmetry between positive and negative stimuli in the ability to sustain the observing response (Auge, 1974; Dinsmoor, Browne, Lawrence, 1972; Lieberman, 1972; Mueller & Dinsmoor, 1984).

1.2. The Choice Procedure

The observing studies can be divided into two groups. In the first group, are the experiments using Wyckoff's procedure: animals are exposed to a mixed schedule of reinforcement, usually with reinforced and non-reinforced components, and a specific response (pressing a pedal but, more commonly, pecking a key) changes the schedule to a multiple schedule, wherein a discriminative stimulus signals which component is in effect. In this procedure, the default state is not having the discriminative stimulus freely available though it can be obtained via a specific response. The experiments using this procedure are usually interested in knowing which variables influence the emission of the observing response.

The second group of studies are designed to answer the question of whether animals *prefer* to have discriminative stimuli. In this procedure, there is no default state and to access the discriminative stimuli the animal has to choose between two options: the Informative and the Non-informative options. In the former, discriminative stimuli are presented informing the subjects about the schedule currently in effect, whereas in the latter stimuli are also presented but they are uncorrelated with the schedule in effect. Prokasy (1956), for instance, placed rats in the middle arm of an E-maze and studied whether they would go to the left or right arm. After choosing one of the arms, rats had to wait 30 s to enter the goal box, that could be baited or not (with both arms baited half of the trials). Even though the probability of reinforcement was the same in both arms, the rats preferred the arm that was white when baited and black when not baited over the arm whose color was uncorrelated with the presence of food. Using pigeons as subjects, Bower, McLean and Meacham (1966) found the same result. When pigeons were given a choice between an Informative Option with signaled outcomes and a Non-informative Option with unsignaled outcomes, animals preferred almost exclusively the Informative Option (see also Green & Rachlin, 1977; Roper & Zentall, 1999).

2. The Suboptimal Choice Task

In all aforementioned experiments, the probability of reinforcement was the same whether animals chose the Informative or the Non-informative Option, so there was no cost associated with preferring information. Resorting to a concurrent-chains procedure, Kendall (1974) asked whether animals would be willing to lose food in order to obtain information about forthcoming reward or no reward. Because this is also the procedure used in the present dissertation, we describe it in greater detail (Figure 1).

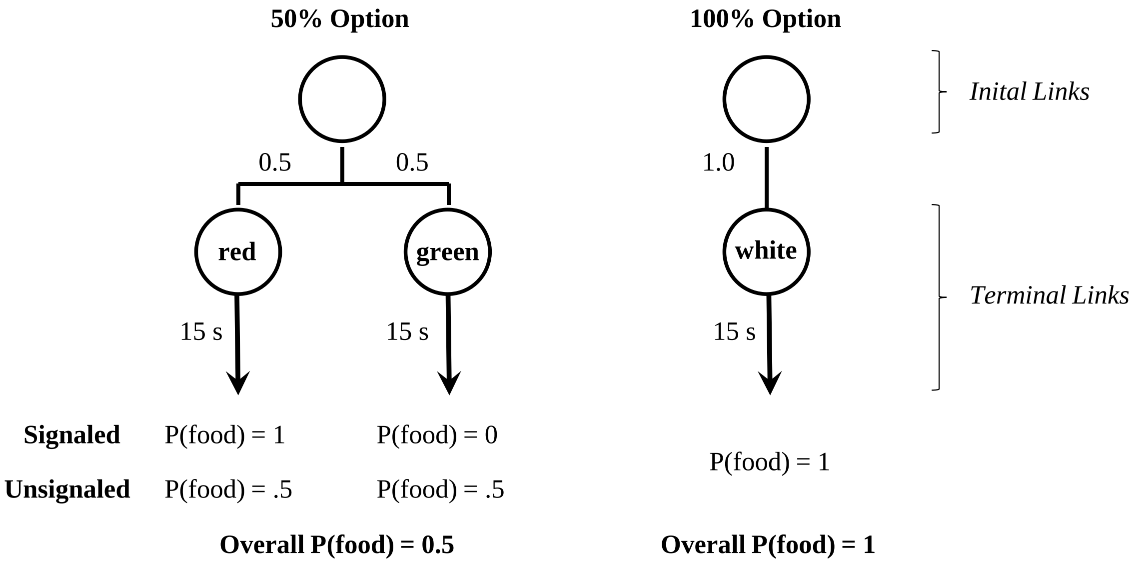


Figure 1. Concurrent-chains procedure used by Kendall (1974).

A concurrent-chains schedule is composed of two links: the initial link (choice phase) and the terminal link (outcome phase). During the initial link, the pigeon has two keys concurrently available and choice of one of the keys turns the other key inoperable (usually the unchosen key turns off), and the schedule proceeds to the terminal link in the chosen key. From the initial to the terminal link, the key color usually changes and stays on for a certain duration. The terminal link may end with or without food.

In Kendall's (1974) study, pigeons were divided into two groups: the Signaled Group and the Unsignaled Group. In the Signaled Group, animals chose between two options: the 50% and the 100% options. If they chose the 50% Option, on half of the trials the key turned red for 15 s and always ended with food (a "good-news" stimulus); on the other half of these trials, the key turned green for 15 s and always ended without food (a "bad-news" stimulus). If they chose the 100% Option instead, the key always turned white and was always followed by food. For the Unsignaled Group the procedure was the same, with the exception that the red and the green keys were both followed by food on half of the trials. Kendall found that the Unsignaled birds preferred the 100% Option, but most of the Signaled birds preferred the 50% Option. This surprising result was interpreted as evidence that the reduction of uncertainty was reinforcing (*uncertainty-reduction hypothesis*). More specifically, for both groups the initial link of the 50% Option predicted food on 50% of the trials. However, in the Signaled Group, when the 50% Option was chosen, the terminal link stimuli eliminated the uncertainty about food delivery, because they were perfectly correlated with food or no food.

On the other hand, in the Unsignaled Group, the terminal link of the 50% option did not reduce the uncertainty about food. If reduction of uncertainty is reinforcing, then animals should show at least a bias for the Informative Option even though it provides less food on average.

The original Kendall's (1974) findings were later challenged by Fantino, Dunn and Meck (1979) that found a possible confound in Kendall's procedure. Still today it is not clear whether animals reliably prefer the 50%, leaner Option with a signaled procedure: Some studies report results consistent with Kendall's original findings (Belke & Spetch, 1994; Experiment 3 of Dunn and Spetch, 1990; Kendall, 1974, 1985), whereas others question its reliability (Experiment 2 of Dunn and Spetch, 1990; Fantino et al., 1979; McDevitt, Spetch, & Dunn, 1997; Smith & Zentall, 2016; Spetch, Belke, Barnet, Dunn, and Pierce, 1990). Given that Spetch's lab has found some evidence in favor of a preference for the 50% Option, Dunn and Spetch (1990) proposed a mechanism to explain this preference based on *delay-reduction theory* (Fantino, 1969). Whereas the *uncertainty-reduction hypothesis* proposes that the 50% Option has more value than the alternative because it supplemented information about the trial outcome (food vs. no food), the Dunn and Spetch's hypothesis proposes an improvement in terms of the average delay to food. This hypothesis will not be further explored here because, as the *uncertainty-reduction hypothesis*, it clearly predicts a preference for the 50% Option and the evidence for this result is not strong. Independently of the models, and focusing on empirical data, even indifference between the two options would warrant further exploration, because maximization of food intake predicts an absolute preference for the 100% Option.

Given the inconclusive nature of the above-mentioned results, Gipson, Alessandri, Miller, and Zentall (2009) tested whether animals would prefer the Informative Option when the difference in the overall probability of reward between the two options was reduced (from 50 to 25%). With this aim, they gave pigeons a choice between an Informative Option that provided signaled rewards in 50% of the trials and a Non-informative Option that provided unsignaled rewards in 75% of the trials (Figure 2, left panel). With these parameters animals showed a stronger (around 69%) and more reliable preference for the Informative Option (13 of the 16 pigeons preferred this option).

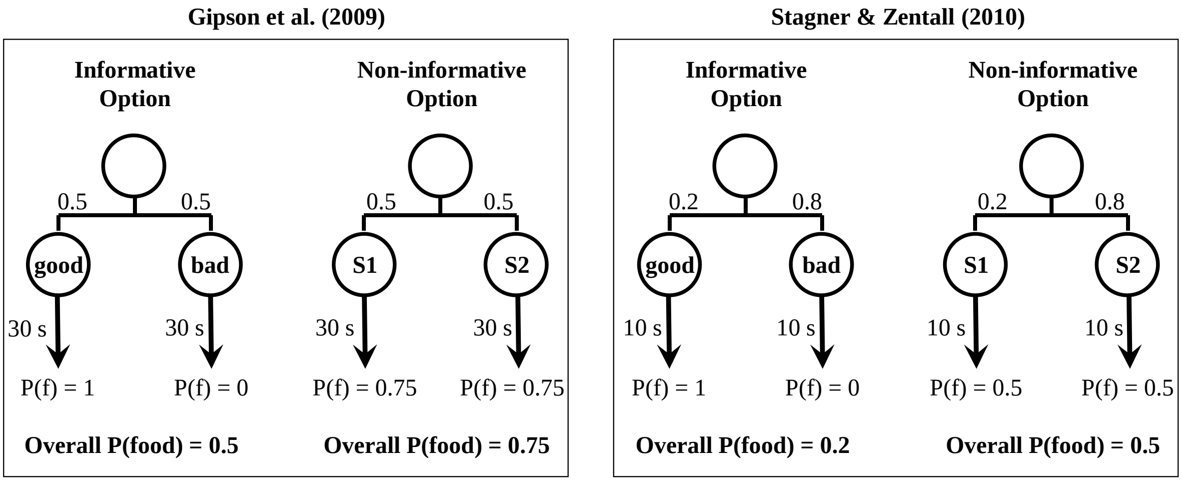


Figure 2. Schematic of the design used in Gipson et al. (2009), left panel, and in Stagner and Zentall (2010), right panel.

An even stronger effect was found when Stagner and Zentall (2010) decreased the overall probability of reinforcement in *both options* (20% in the Informative Option vs. 50% in the Non-informative Option, right panel of Figure 2). In their experiment, preference for the Informative Option reached 97%. How could such a strong and 'irrational' preference be explained?

2.1. Contrast effect?

When Stagner and Zentall (2010) found a strong suboptimal choice, they interpreted the results in terms of contrast. Specifically, in the Informative Option animals expect reward on 20% of the trials, but when the good-news stimulus appears, the expectancy abruptly increases to 100% (an 80% increase); if the bad-news stimulus appears instead, there is a small decrease in reward expectancy (from 20 to 0%). On the other hand, expectancy in the Non-informative Option remains unchanged when the terminal stimuli are presented (the animal initially expects reward on 50% of the trials and when either of the terminal stimuli is presented this value remains the same). Even though the bad-news stimulus is much more frequent that the good-news one, the high positive contrast produced by the latter presumably enhances the value of Informative Option relative to the other option. This interpretation is also consistent with the stronger suboptimal preference found in Stagner and Zentall (2010), where the positive contrast was of 80%, as compared with the preference found in Gipson et al. (2009) where the positive contrast was only of about 50% (Figure 2).

This explanation, however, could not account for later results. For example, Stagner, Laude, and Zentall (2012) gave pigeons a choice between two options, both with discriminative stimuli (good and bad news). In one of the options, the good-news stimulus was shown on 20% of the trials and on the other option it was shown on 50% of the trials. If contrast between the expected probability of reward in the initial and terminal links was responsible for suboptimal preference, animals should prefer the 20% option because it yielded a greater contrast (80% vs. 50%). Contrary to this prediction, animals were indifferent between the options. This result also suggests that the probability of the terminal-link signals and, as a consequence, the overall probability of reinforcement do not seem to influence choice.

2.2. Uncertainty aversion?

A different hypothesis to explain suboptimal choice proposes that animals prefer the Informative Option because it allows them to avoid the uncertainty associated with the Non-informative Option. To test this hypothesis, Zentall and Stagner (2011a) eliminated the uncertainty in this option by manipulating the amount of reward after each signal instead of the probability of reward. Figure 3 shows their design.

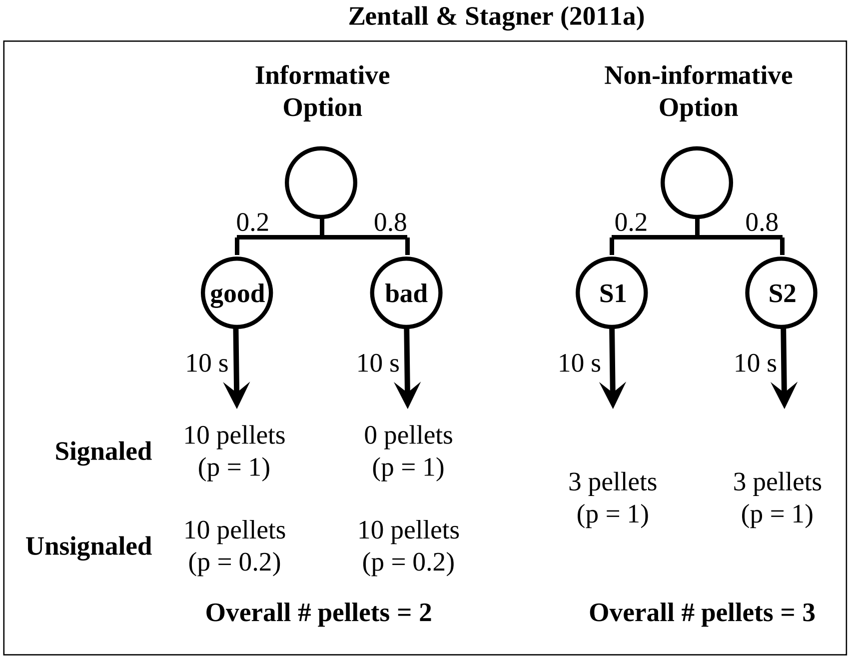
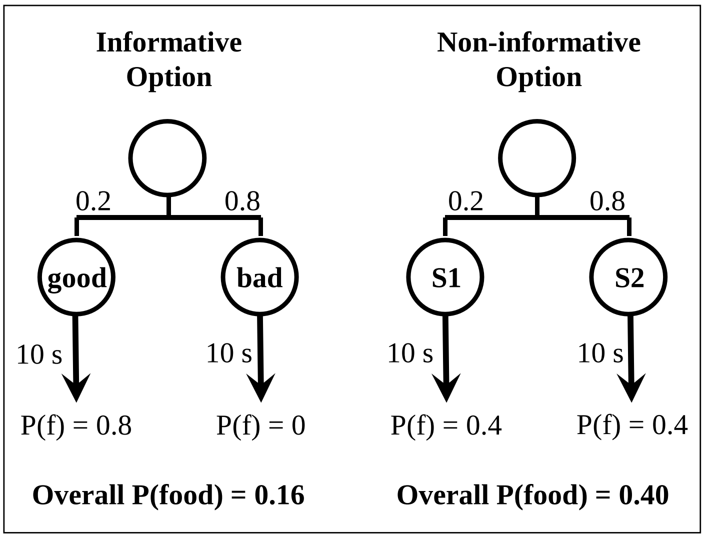


Figure 3. Schematic of the design used in Zentall and Stagner (2011a).

In the Signaled Group, the Informative Option delivered 10 or 0 signaled pellets, and the Non-informative Option always delivered 3 pellets. Even though the Non-informative Option provided more food on average (3 pellets *vs.* 2 pellets) and its outcome was predictable, animals showed an 82% preference for the Informative Option. Yet, because the overall amount of food was similar between options, it was possible that animals did not discriminate the amounts and the result due to some preexisting bias. However, when in the Unsignaled Group the stimuli in the Informative Option were made nondiscriminative, animals chose the 3-pellet option in 80% of the time. These results show that suboptimal preference is not a result of uncertainty aversion, and that the discriminative stimuli in the Informative Option play a pivotal role (Laude, Beckman, Daniels, & Zentall, 2014; Laude, Stagner, & Zentall, 2014; Zentall & Stagner, 2011a).

2.3. Certainty attractiveness?

It remains possible that animals are attracted to the Informative Option because it is the only alternative that provides a signal that is perfectly correlated with food – the good-news stimulus – and the value of this stimulus may be overweighed (see the certainty effect, Tversky & Kahneman, 1986). To test this hypothesis, Zentall and Stagner (2011b) compared performance of a group with the design of Stagner and Zentall (2010, right panel of Figure 2) with a different group in which the good-news stimulus was not a perfect predictor of food. Figure 4 shows the design for that group.



Zentall & Stagner (2011b)

Figure 4. Schematic of the design used in Zentall and Stagner (2011b).

Zentall and Stagner (2011b) decreased the probability of food following the good-news stimulus to 80%. Also, to maintain the overall probability of reinforcement in the Non-informative Option 2.5 times higher than in the Informative Option (as in Stagner & Zentall, 2010), the probability of food following the non-informative signals was decreased to 40%. Even though the Informative Option now provided an unreliable signal, animals still strongly preferred this option (> 90%).

2.4. Bad-news avoidance?

After Zentall & Stagner's (2011b) experiment and after rejecting several possible explanations for this suboptimal preference, there was enough evidence supporting the idea that animals were choosing solely based on the value of the conditioned reinforcers, that is, the stimuli associated with food delivery (thus excluding the bad-news stimulus from the 'equation'). More specifically, the hypothesis was that choice was based not on the objective probabilities of food conveyed by the initial links (e.g., 20 *vs.* 50% in the procedure used by Stagner and Zentall, 2010), but on the probabilities of food signaled by the terminal links acting as conditioned reinforcers (100% in the Informative Option *vs.* 50% in the Non-informative Option). By comparing these values, one can correctly predict which option animals choose in all experiments described so far. Moreover, this analysis also explains why the preference for the Informative Option found in Gipson et al. (2009) was weaker than that reported by Stagner and Zentall (2010, compare panels of Figure 2): In the first case the discrimination was between 100% and 75% whereas in the second it was between 100% and 50%. Also, in Kendall's (1974) design (Figure 1), animals in the Signaled Group should be indifferent between the two options (Experiment 2 of Dunn and Spetch, 1990; Fantino et al., 1979; McDevitt, Spetch, & Dunn, 1997; Smith & Zentall, 2016; Spetch, Belke, Barnet, Dunn, and Pierce, 1990). In brief, this hypothesis assumes that: (a) animals do not take into account the bad-news stimulus because it is not a conditioned reinforcer; (b) the probability of the good- and bad-news stimulus should not affect choice, and; (c) preference arises from the comparison between the probability of food in the good-news stimulus and the probability of food in the non-informative stimuli.

Focusing on the first assumption, Stagner, Laude, and Zentall (2011) tested whether the bad-news stimulus was not taken into account because animals avoided it, for instance, by turning their back to it. For half of the birds the bad-news stimulus was a houselight, a diffuse light that animals would hardly not observe; for the other half the good-news stimulus was the houselight (and the bad-news stimulus was a keylight as usual). The groups did not differ in preference, and as usual preferred the Informative Option. This suggests that simple avoidance of the bad-news stimulus cannot explain why this stimulus does not seem to affect the overall value of the Informative Option.

Later, Laude et al. (2014) showed that the bad-news stimulus is an inhibitor at the beginning of the experiment, but it loses inhibitory value as training progresses. To test for inhibition, they used the compound-cue test (or summation test). The bad-news stimulus was a white bar on a black background and the good-news stimulus was a red hue. Early and later in training pigeons were tested with the good-bad compound stimulus: a red bar on a black background. If the bad-news is an inhibitor, the response rate on the compound stimulus should be substantial lower than in the good-news stimulus alone. In the present task, it is relatively common for animals to start choosing the optimal option and later start preferring the suboptimal one. Early in training, when animals were still preferring the optimal, Non-informative Option, animals responded significantly less in the compound stimulus. However, later in training when animals were reliably preferring the suboptimal, Informative Option there was not a significant decrease in the response rate in the compound stimulus. This strongly suggests that the bad-news stimulus is not an inhibitor and it does not devalue the Informative Option.

Despite several differences in procedure and in purpose, Mazur (1989, 1991, 1993, 1995, 1996, 1998) also found evidence that the value of an option is only influenced by the time spent in the presence of conditioned reinforcers. Mazur was interested in which stimuli influence the value of an option that delivers delayed rewards, and resorted to the adjusting-delay procedure (Mazur, 1987). With this procedure, Mazur found that, when an option produces with probability p a good-news stimulus for d seconds, and with probability $1-p$ a bad-news stimulus also for d seconds, animals are indifferent between this option and another option that always delivers food after d seconds, as if the bad-news stimulus did not influence the value of the first option (Mazur, 1996).

Thus far, the results suggest that the bad-news stimulus plays little to no role in this choice task. If this assertion is true, then manipulations related to this stimulus should not affect choice. More specifically, increasing for instance the salience, duration or probability of the bad-news stimulus should not affect the value of the Informative Option and thus preference. However, the procedures, designs, and parameters of the experiments have widely varied, and a more systematic approach to test the influence of the bad-news stimulus on choice is needed.

3. The present dissertation

The main goal of the present thesis was to investigate the role of the bad-news stimulus on choice with pigeons. To accomplish this goal, we established two objectives, one more theoretical and the other more empirical.

First, at a theoretical level, we were interested in testing an optimal foraging model, the Reinforcement Rate Model (RRM; Vasconcelos, Monteiro, & Kacelnik, 2015), to explain the typical results observed in this task. We took this approach because, as animals behave (objectively) suboptimally in this task, this behavior may seem to question the foundation stone of optimal foraging theory itself (McDevitt, Dunn, Spetch, & Ludvig, 2016; Zentall, 2016). However, as will become clear along the present dissertation, not only this suboptimal behavior does not contradict optimal foraging theory, as it can, at least partially, be explained by it. In Chapter II we briefly introduce optimal foraging theory. In Chapters III to V we describe six experiments in which we use the Vasconcelos et al.'s (2015) model to explain the results. In some experiments some adaptations to the model were necessary, and for that we also resorted to optimal foraging principles. Moreover, whenever possible, we compared the predictions of the RRM with the Mazur's (1984, 1987) Hyperbolic Discounting Model (HDM). As the RRM assumes long-term maximization and the HDM assumes short-term maximization, comparing the predictions of these models allows us to better test the long-term maximization assumption of optimal foraging models. As stated before, a third model that has been used to explain performance in this task. The Dunn and Spetch's (1990) hypothesis will be briefly described in Chapter III of the present dissertation, but as it clearly predicts a preference for a 50% Option *vs.* 100% Option (see Figure 1), and that result does not seem to be neither strong, nor robust, we decided not to explore it further. Instead, we resorted to the two quantitative models that until now have accounted reasonably well for the data reported in the literature: the RRM and the HDM.

Second, at an empirical level, we wanted to systemically test the hypothesis that presenting a terminal stimulus that is never followed by food does not affect choice. Even though some experiments have indicated that this was the case, in our view the previous procedures could be improved to obtain clearer data. For this goal, we used Stagner and Zentall's (2010) design (right panel of Figure 2). In Chapter III we present two experiments. In the first experiment, we manipulated the frequency of the bad-news stimulus, from 80 to 100%, while keeping the probability of reinforcement in the Non-informative Option at 50%. In the second experiment, we manipulated the duration of the bad-news stimulus with a titration procedure: if pigeons preferred the Informative Option we increased this duration and if pigeons preferred the Non-informative Option we decreased it. To truly know a phenomenon, we need to know its boundary conditions. In these two experiments, by either using an extreme probability or an extreme duration of the of bad-news stimulus, we aimed at testing the conditions under which pigeons do not take into account the bad news and choose suboptimally.

In Chapter IV, we further explored the notion that animals do not take into account the bad-news stimulus. We had shown that this stimulus does not influence the value (or at least the choice) of the Informative Option. However, do animals process this stimulus but simply do not associate it with the choice of the Informative Option or is this stimulus not processed at all, and thus, for all intents and purposes a never-presented stimulus? Even though Vasconcelos et al.'s (2015) RRM was not explicit in this regard, its assumptions were consistent with the prediction that, if animals could peck a key to reject the bad-news stimulus, they should do so. This implicitly states that the bad-news stimulus is processed because, otherwise, why would animals peck a key to reject something that does not exist? In two experiments, we used a modified version of a sequential-encounters procedure (Lea, 1979), in which we gave pigeons the option of rejecting the presented stimulus. We manipulated the bad-news duration to see if it had an effect on performance and also manipulated the intertrial duration because it has been shown to affect performance in sequential tasks.

In the two experiments of Chapter V, we asked what animals would do if, the bad-news stimulus turned into a "not-so-bad" stimulus. On the one hand, we were interested in finding how manipulating the probability of reinforcement following this stimulus affected preference for the Informative Option. On the other hand, we were concerned with how animals dealt with the duration of this stimulus. If a stimulus that is *never* followed by food is not taken into account, but a stimulus that is *always* followed by food is taken into account, what happens when the stimulus is *sometimes* followed by food? To address this question we explored how the probability of attending to a stimulus and the probability of that stimulus being followed by food are related to each other.

Finally, Chapter VI summarizes the obtained results.

CHAPTER II [1]

ON THE STRUCTURE AND ROLE OF OPTIMALITY MODELS IN THE STUDY OF BEHAVIOR

---

[1] This Chapter reproduces the publication:

Vasconcelos, M., Fortes, I., & Kacelnik, A. (in press). On the structure and role of optimality models in the study of behavior. *APA Handbook of Comparative Psychology*. Washington, DC: American Psychological Association

1. Introduction

The use of optimality models in behavioral biology and comparative cognition stems from the view that the mind is a product of the same evolutionary process that leads to kidneys, wings, petals, eyes, or monkeys' tails. The chief participant in this process is natural selection, a mechanism identified in its fundamentals by Charles Darwin and Alfred Russel Wallace towards the middle of the 19th century. Its essence is that randomly originated, heritable variation inevitably leads to evolutionary changes, because traits that best promote reproduction and survival increase their representation in the species across generation and thus progressively become the species' norm. In the long-term, biological traits, including psychological mechanisms, appear as if they had been designed to maximize reproductive success, and mathematical models that assume optimal, fitness-maximizing design can be used to generate testable hypotheses about decision mechanisms.

Optimality techniques are common in many areas of biology, as when anatomists examine the shape of wings as being designed either for flapping or gliding flight. George Williams (1966) highlighted the relevance of assuming fitness-maximizing design to psychology, by asking rhetorically "Is it not reasonable to anticipate that our understanding of the human mind would be greatly aided by knowing the purpose for which it was designed?" (p. 16). Williams –like us- uses the term design in a non-teleological, process-based way, and not in the contorted irrational sense that is predicated by supporters of the "intelligent design" idea. In Williams' approach, the mind's design "purpose" (on which optimality techniques rely) is different from the goals driving the behavior of the organism in which the mind is embodied. The mind's biological purpose *sensu* Williams can be described as the "goal" or directionality of natural selection, not the goals of the acting agent. For instance, mating behavior may be driven by an organism's pursuing of sexual desire, but the purpose to which biological optimality refers is reproductive output, with desire being the tool through which the organism's behavior is manipulated by its genes. Similarly, fear, hunger, aggression, or maternal love are all tools of natural selection to make individuals "do the right thing", namely maximize their representation in future generations. Thus, agents' motivations appear as designed for a purpose, namely to maximize fitness. Those ancestors whose mental traits we inherited were not necessarily the happiest, most emotionally balanced, or most intelligent ones, but those whose minds led to maximize the recruitment of descendants to the breeding population.

Optimality is thus a framework for devising hypotheses about how animals work and what is important to selection, and our overall message is that this framework may be very useful to psychologists, provided it is properly understood and used in conjunction with empirical research. To make the optimality work it is especially important to avoid several trivial but frequent misrepresentations, namely that evolutionary biologists expect animals to behave perfectly, that optimality models require the organism to compute optima in order to act, and that the hypothesis under test is whether animals are optimal. The assumption of those using optimality techniques is not that organisms make perfect fitness-maximizing choices in all circumstances, but that their psychological mechanisms reflect those of ancestors that outperformed their contemporaneous conspecifics. Inasmuch as present circumstances reflect the species' past, such mechanisms are expected to cause decisions that are, on average, adaptive for present members of the species. One way to dispel such misinterpretations is to focus on the structure of optimality models and to follow some examples in detail. This is our objective in the present article, and we'll pursue it by illustration, describing how the optimality approach has aided research in a few clear examples.

1.1. Components of optimality models

There are alternative ways to carve up models' components, but here we follow the approach of Kacelnik and Cuthill (1987). They argued that optimality models are an assemblage of at least three interconnected assumptions, all of which are in turn hypotheses, since they are independently testable: the *Strategy Set*, the *Feedback Function,* and the *Currency*.

1.1.1. The strategy set.

For the purpose of modeling, optimality practitioners see all behavior as choice, but this does not mean conscious deliberations. The choice may be between discrete alternatives such as different food types (e.g., Pulliam, 1974), mates (e.g., Slagsvold & Dale, 1991), or forms of locomotion (say walking or flying; e.g., Bautista, Tinbergen, & Kacelnik, 2001), or between points along a continuum, such as intensity of parental effort (e.g., Wright & Cuthill, 1990), flying velocity (e.g., Houston, 1986), or size of an ejaculate (e.g., Parker & Pizzari, 2010). This means that right from the start, models define the range of potential behaviors, or strategy set. The strategy set used in a particular model is then a hypothesis inspired by observing the organism in its normal circumstances, as the range of alternatives must be realistic (i.e., when studying decision making in horses one can compare trotting with galloping, and when studying decision making in starlings one can compare walking with flying, but species and choices cannot be swapped). This is not as trivial an issue as it may seem, because while anatomical constraints may be obvious, psychological ones are not. For instance, shiny cowbirds (*Molothrus bonariensis*) in the Americas parasitize the reproductive effort of other species, and their offspring share the nest with those of their hosts (Gloag, Fiorini, Reboreda, & Kacelnik, 2011). Hosts seem unable to feed preferentially their own offspring, probably due to psychological mechanisms evolved to drive parental behavior in unparasitized nests, and so the strategy set for some optimality models of host behavior includes a choice between abandoning the present brood or not, but not the choice of which nestling to feed. These constraints introduce elements of circularity, because a model can only produce as an output a member of its hypothetical strategy set. Although it is important to be aware of this issue, this is not a flaw of the optimality approach, because as we have said from the start, the strategy set is a hypothesis, and thus is itself subject to recursive testing and improvement.

1.1.2. The feedback function.

Each action within the strategy set would have different consequences as a function of the state of the organism and its environment. For instance, a threatened zebra choosing to gallop rather than walk increases its probability of escaping from a stalking lion, but also increases metabolic rate and interrupts grazing, so that the relative payoff of each behavior is not only affected by the probability of escape but also by how hungry is the animal and how easy it is to find food. The feedback function in a model dealing with choice of mode of locomotion would describe probability of escape and/or nutritional consequences as a function of mode of locomotion. The feedback function is thus a hypothesis about what happens to the actor as a function of what it does, and it is meant to be refined with accumulated knowledge. To create a new model, the modeler makes informed guesses to include a manageable number of state-dependent action-consequence relations considered to have influenced the evolution of the organism’s psychology.

1.1.3. The Currency.

Optimality models are predominantly normative rather than descriptive, because decisions are predicted as if the actor intends to maximize Darwinian fitness, but they are not just normative, as they cannot avoid including assumptions about mechanisms of behavior. This is in part because Darwinian fitness, defined as the contribution to the species' gene pool in later generations, is not measurable at the time scale of behavioral studies. Each action may have a minute effect, but the pattern of choice integrated along a lifetime and across generations may impact the fitness of individuals and of the alleles (variants of genes) that they carry. To build optimality models that are testable at a behavioral time scale, the modeler identifies short-term, measurable variables that are good candidates to have a clear relation to long-term fitness. Examples of such variables are rate of intake, vulnerability to predation, probability of avoiding starvation, or balance between nutrients. As with the other two categories of model components, these currencies are hypotheses: if an animal does not act so as to maximize a candidate currency, it may be that that particular variable is not a significant bottleneck with respect to natural selection, and new models will modify the currency assumption.

1.2. Predictions of optimality models

Optimality models are used to predict or explain behavior in natural and experimental circumstances, and to design experimental protocols that challenge those predictions. To predict behavior, the modeler asks which member(s) of the strategy set maximizes the currency given the feedback function. It is mainly in the interaction with empirical data that the value of this research program has to be judged. Like all scientific ideas, the predictions of optimality models sometimes succeed and sometimes fail. Confirmation of the predictions is valuable if the predictions supported are novel, previously unexplained, and to some degree counterintuitive. Failed predictions are just as informative (often even more) as those corroborated. If a model's prediction is falsified, at least one of the component hypotheses must be wrong, and a new model needs to be formulated after revision. Once again, the method does not intend or allow for corroboration or falsification of claims that animals are optimal. The assumption that natural selection is an optimizing process plays a metatheoretical role for biologically inspired behavioral science: it is not tested, but it underlies its validity. Optimality models thus help to shift behavioral sciences from a descriptive towards a hypothetico-deductive approach, and increase the precision and explicitness of our reasoning.

We will not review optimal foraging models exhaustively (for detailed treatments see, for example, Houston & McNamara, 1999; Kacelnik & El Mouden, 2013; Stephens, Brown, & Ydenberg, 2007; Stephens & Krebs, 1986). Instead, we aim at showing with examples from our own work, predominantly using European starlings (*Sturnus vulgaris*) as experimental systems, that the functional rationale underlying research in behavioral ecology is a powerful tool, sometimes underestimated by experimental psychologists (just as behavioral ecologists often overlook the importance of mechanisms). In our view, the optimality approach complements rather than competes with the mechanistic approach prevalent in the psychological sciences. We believe that cross-fertilization between evolutionary biology and animal psychology is the best way to fulfill Tinbergen's (1963) desideratum of a behavioral science straddling across mechanisms, adaptive value, ontogeny and phylogeny.

2. Patch exploitation. The Marginal Value Theorem

Economists have long been aware that benefit often increases less than linearly with effort invested (Samuelson, 1937). For instance, a baker may sell more loaves by increasing his opening hours, but not at a constant rate: as a larger fraction of neighborhood residents have bought tomorrow's bread, the additional loaves sold in yet another opening hour (the marginal benefit) decline, and at some point the baker benefits more by allocating his time to other sources of revenue, such as making cakes instead. The analysis of strategies for switching between activities has received much attention in foraging contexts, and one of the best-studied optimal foraging models is known as the Marginal Value Theorem (MVT; Charnov, 1976b; Parker & Stuart, 1976). One specific scenario where the MVT applies is in Central Place Foraging, as embodied by a bird that provisions dependent nestlings. Brood provisioning is amenable to modeling because it can be decomposed in regular cycles: the parent leaves the nest, lands at a foraging patch, gathers a certain amount of food (a fraction of which it eats), and at some point flies back to deliver a load of food to its brood. Over the day, it repeats this cycle hundreds of times, and virtually all of the daylight hours are occupied flying to and from the foraging patches, collecting food, and delivering it to the nestlings. As discussed above, to formulate a model one has to decide on a currency, a strategy set and a feedback function. We now see how this works in the present case.

Since food provisioning affects the chicks' fitness, the rate at which food is delivered is a sensible a priori candidate as a model's currency. As for the strategy set, one could consider all actions that are under the bird's control, including allocation of captures between the parent and the brood, the time when to stop collecting food to fly back towards the nest, or distribution of food among the nestlings. Different models and experimental studies address each of these decisions (e.g., Kacelnik, Cotton, Stirling, & Wright, 1995; Kacelnik & Cuthill, 1990), but here we focus on when to stop gathering food. In this example the strategy set is defined by the range of potential times in the patch, which is a continuous variable, and the feedback function is how the currency (provisioning rate) varies as a function of time in the patch. The analogy with the baker's example arises from the fact that when birds collect multiple prey in their beaks in each trip, load does not increase linearly as a function of patch time, because prey already held slow the bird down. The MVT then applies in a very straightforward manner, as follows.

The maximized currency (provisioning rate, R(t)) can be expressed by

$$R(t) = \frac{G(t)}{\tau + t} \tag{1}$$

where G(t) is the gain curve, expressing the load accumulated as a function of time since arrival at a patch, τ is the mean travel time for round trips between nest and feeding patch, and *t* is the time between landing in the patch and taking off (called patch time below). The problem is to find the value of *t* that maximizes R(t) given the shape of G(t) and the value of τ. If it is known that capture rate decreases with time in the patch (i.e., the 2nd derivative of G(t) is negative), then we know from calculus that the optimal *t* is the point at which the first derivative of R respect to *t* is null[2], provided that at that point the 2nd derivative is negative. This value, t_{op}, is the predicted patch time. Since patch time and load are directly related through G(t), predicting t_{op}, also specifies the optimal load per trip, $G(t_{op})$.

Figure 1a depicts the problem graphically, plotting lines passing through a notional start of a foraging cycle and loads at different possible departure times. The slope of these lines is given by the ratio of total gain G(t) over total cycle duration (τ + *t*), which is precisely R(t), the currency we want to maximize. The steepest of these lines has slope $R(t_{op})$ and is

---

[2] The first derivative of R(t) is given by $\frac{dR}{dt} = \frac{G'(t)(\tau+t)-G(t)}{(\tau+t)^2}$, which is zero when $G'(t)(\tau + t) = G(t)$ or $G'(t) = \frac{G(t)}{(\tau+t)}$ namely when the first derivative of G(t) equals the overall rate of provisioning.

tangent to G(t). At this point, the first derivative of G(t) equals R, the overall rate of provisioning taking into account travel time as well as the shape of G(t) and *t*. In this example, the first derivative of G(t) is a monotonically decreasing function of *t*, while R (which is what we aim at maximizing) has a peak at t_{op}.

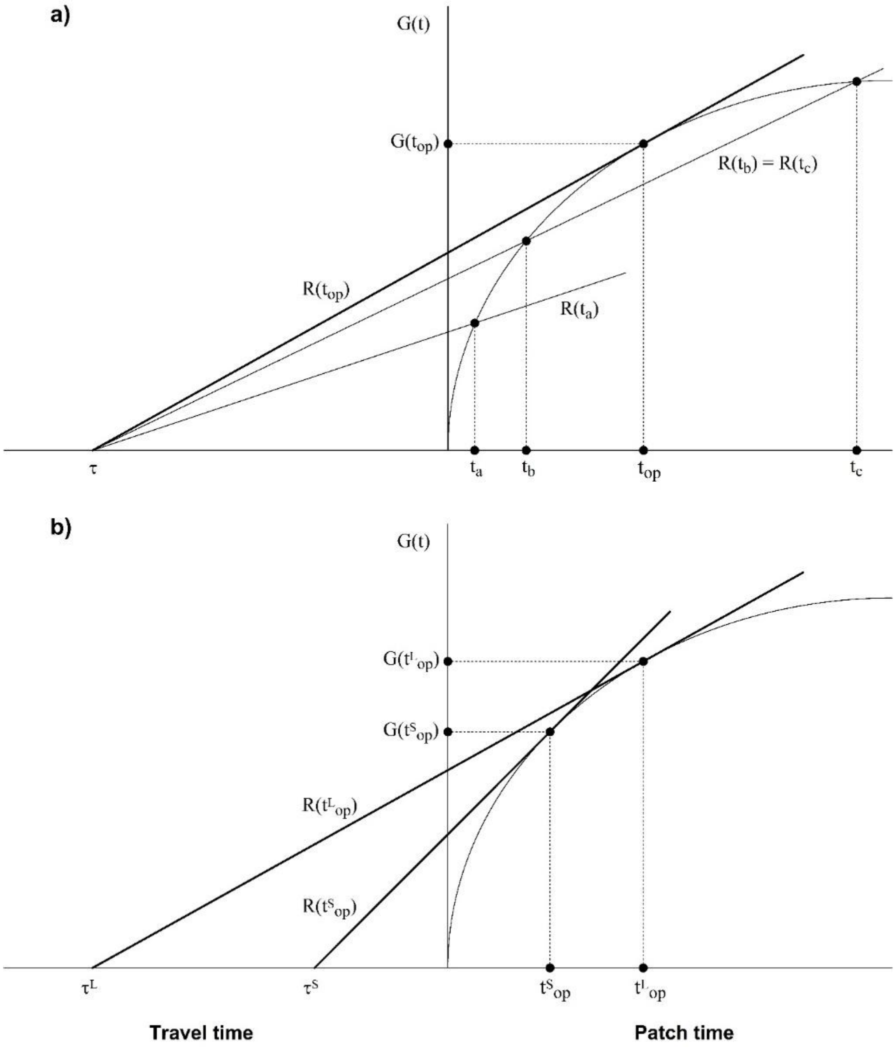


Figure 1. Graphical representation of the marginal value theorem as applied to a central place forager. The origin of coordinates is set at the arrival time to a patch. The horizontal axis shows the total travel time (round trip) growing to the left and patch time growing to the right; the vertical axis shows accumulated food, with the curvilinear function showing total gains as a function of patch time. An animal leaving all patches after collecting food for a time t will experience an overall rate of returns R, given by the ratio of accumulated gains to the sum of travel and patch times. In (a) several potential leaving times are shown. The line with slope R(top) represents the maximum rate of prey acquisition, and serves to identify the optimal patch leaving time. Also shown for comparison are two alternative strategies, with rates R(ta) and R(tb) = R(tc), both less profitable than R(top); (b) shows the effect of travel time. When the round trip travel decreases so does the optimal patch residence time and

consequently the optimal load size. Notice that only at the optimal departure time is the slope of G(t) identical to the overall rate R.

According to the model, the rate-maximizing policy is obtained by adhering to a (mathematically) simple rule: stay in the patch as long as the local rate (G'(t), the first derivative of G(t)) exceeds the expected overall rate R. This rule is prospective, since the ideal decision maker simply chooses where to invest its next unit of time. The strategy set in the case of this model includes the capacity to tune behavior to the overall rate R in the environment, either by learning from experience (McNamara & Houston, 1985) or by reading some environmental clue. It also includes the capacity to directly perceive the local rate G(t) as it drops as a function of patch time. The mutual dependence between R and t_{op} is solved through convergence until experienced and expected R are equal.

Figure 1b shows how travel time, τ, affects optimal patch time. As τ increases from τ^S to τ^L, the optimal residence time and consequently the optimal load increase. This is intuitive because the less time spent traveling the higher is the overall rate of return R, and G(t) drops to R earlier. To our knowledge, the prediction that longer travel time should lead to longer patch times was met in every published experimental test of the MVT.

The MVT as described so far incorporates simplifications, including the following:

- G(t) is a continuous function, but foraging animals encounter discrete prey items, so that responding according to how its slope declines with patch time may pose implementation problems.
- Foraging cycles are assumed to be identical, and parameters are entered in the model only as averages, but in practice patches differ from each other as do travel times to them, so that variance, as well as averages, may have an impact.
- The currency R is provisioning rate, but parents must engage in other activities such as territorial defense or anti-predator behavior, and this may impose tradeoffs.
- The model deals just with maximizing provisioning rate, but the parent needs to eat to stay alive.
- The currency as discussed so far does not take into account known metabolic costs of foraging and flying.
- The model is mute regarding mechanism: the agent does not need to compute the optimum as the modeler does, but it must acquire and process the relevant information somehow.

Exposing these simplifications is an important contribution of the model. All of these simplifications are amenable to theoretical refinements and experimentation, and have been dealt with in the foraging literature. One of us (Kacelnik, 1984), for example, applied the MVT to European starlings feeding nestlings. Starling pairs usually make between them in the order of 400 foraging trips per day, bringing up to eight prey per trip. In a field experiment, starlings learned to collect mealworms from an artificial patch that was reset in every visit, and where G(t) was implemented by delivering worms at increasing intervals. The delivery rule was a discrete approximation to $\mathrm{I}_i = e^{(i/1.52)}$ where I_i is the time between landing and delivery of the i_{th} prey in each visit, in seconds. As the birds collected prey as they were delivered, they experienced $\mathrm{G(t)} = 1.52 \ln(\mathrm{t})$.

To test the impact of travel time, the distance (travel time) between patch and nests was also manipulated, allowing for *a priori* predictions of the optimal patch time, or equivalently of the optimal load. For the traveling distances tested (up to 1 km from the nest), the observed loads increased with travel time and were close to the predicted optima. However, the model showed a quantitative deviation: it slightly but systematically underpredicted the loads carried by the birds. One of the model’s simplifications turned out to be the culprit: the currency (gross rate of delivery) treated all time components in the cycle as equivalent (i.e. they all caused the same loss of foraging opportunity), but flying time is more energetically costly than foraging on the ground or time spent in the nest. A realistic improvement of the currency hypothesis to include energetic costs (Cowie, 1977; Kacelnik & Houston, 1984) increased the quantitative fit between model and data (Figure 2).

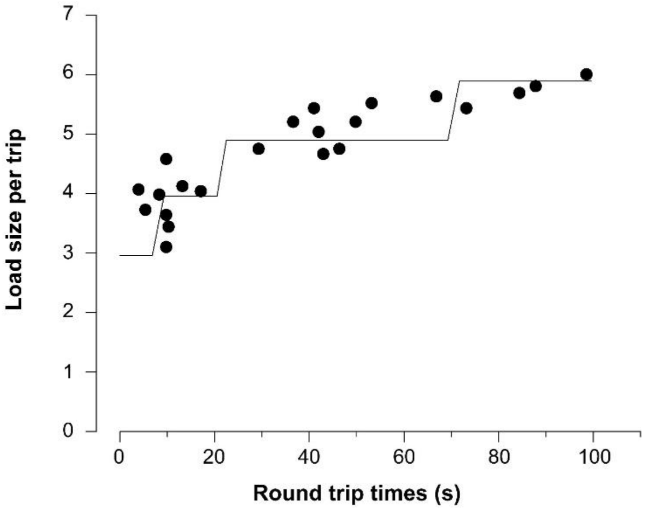


Figure 2. Load size as a function of round trip times in a field experiment with starlings as subjects. Each dot represents the average number of mealworms collected over approximately 50 trips to the

same foraging site; the line represents the predicted optimal number of prey according to the MVT when physiological costs are considered. Adapted from Kacelnik (1984).

In addition to linking foraging behavior to economics, the MVT applies to many other biological problems. One classic example is that of male dung flies (*Scatophaga stercoraria*), that compete for the opportunity to mate with females arriving at cowpats to lay their eggs. Copula in these flies takes variable times for two reasons: because the transfer of sperm increases with copulation time and because males guard females after ejaculation, to prevent them from copulating with other males. Parker (1970; see also Parker, Simmons, Stockley, McChristie, & Charnov, 1999; Parker & Stuart, 1976) has shown that the sperm of a second male copulating with a female fertilizes about 80% of the eggs. So the strategic question here for each male is how long to spend in each copula. The returns of added copulation time diminish rapidly (in terms of the expected proportion of eggs fertilized) and there is also the opportunity cost of encountering other females. The time each male spends guarding the present female until she lays her eggs plus the time to search for a new female (the analogue of travel time in the starling experiment) can be used to predict the time a male should spend copulating with a female, and the predictions work reasonably well (Parker, 1978).

2.1. The role of psychology in optimal foraging

In summary, animals in many cases do follow the predictions of the MVT, but in its simplified versions this does not address the proximate mechanism(s) by which they do. Behavioral ecologists often address this by postulating that animals may follow simple rules, known as *rules-of-thumb*, capable of engendering behaviors close to the predicted optima (e.g., Davies, Krebs, & West, 2012). Many such rules have been proposed, each to deal with particular features of a problem. Examples include *hunting by expectation* (the animal should leave the patch after a given number of captures; e.g., Gibb, 1958, 1962a, 1962b, 1966), *giving-up time* (the animal should leave the patch when the time since the last capture exceeds a given threshold; e.g., Croze, 1970; McNair, 1982), *patch-residence time* (the animal should leave the patch after a certain exploitation time; e.g., Krebs, 1973), and even Bayesian updating rules that can cope well with variability between patches (e.g., R. F. Green, 1980, 1984). These hypothetical rules are close to the heuristics favored by some critics of optimality in the study of human decision making (e.g., Gigerenzer & Goldstein, 1996; Gigerenzer, Hertwig, & Pachur, 2011), and have similar shortcomings. Their main virtue is that they show that animals can approximate optimal behavior without making the same computations as the scientist. For those that support bounded rationality as if it were an alternative to optimality, heuristics illustrate the same point, namely that individuals do not make the computations required to identify optima and instead follow algorithms that converge to sufficiently appropriate behavior under the ecological conditions in which the decision maker lives. But we are concerned with some shortcomings.

First, most of the rules proposed are domain-specific. For instance, a foraging bird may use waiting time between two prey items as a proxy for the reciprocal of G’(t), departing when experiencing a given interval, or a baseball player may catch a ball not by computing the ball’s parabolic trajectory but by running so as to maintain a constant angle of gaze to the ball as it flies, until both converge. We fear that such rules are probably unsuitable for animals inhabiting heterogeneous habitats and facing a multitude of daily tasks. Members of species that face different demands within and across generations would have to use the correct *rule-of-thumb* in each particular situation, which implies the existence of an extensive library of such rules and a mechanism to select the correct one in each case. Second, the approach mostly ignores psychological mechanisms known in the parallel, but highly related field, of animal learning and cognition. Mechanisms such as reinforcement learning (loosely defined as increasing the frequency of actions that work well) may adjust the organism to a very broad class of problems, ranging from rate maximizing patch times to the right movements to catch a flying ball. The adaptive algorithm is then a learning and developmental process, rather than a specific rule of thumb or heuristic for each individual problem.

As stated previously, we argue that optimality models are a framework to integrate functional and mechanistic hypotheses. Just as adaptive function constrains which psychological mechanisms evolve, broad-domain psychological mechanisms determine the nature of the problems each animal solves.

To illustrate this view we turn our attention to how learning theory and psychophysics relate to the foregoing discussion of the MVT. It is clear from the previous discussion that optimal foraging models involve sensitivity to time intervals such as travel time and inter-capture intervals. This by itself suggests that mechanisms known under the heading of *interval timing* (psychological processing of learned intervals typically in the seconds to minutes range) might be directly relevant (see Vasconcelos, Carvalho, & Machado, in press).

This was demonstrated by Brunner, Kacelnik, and Gibbon (1992; see also Kacelnik & Brunner, 2002) using a laboratory task in which starlings travelled between perches to reach a virtual patch, where on arrival started to receive food pellets with fixed inter-capture intervals (fixed-intervals, FIs). The rate of capture did not decline gradually, but after an unpredictable and unsignaled number of deliveries the patch went dead, and the bird could renew the process by “traveling” to a new patch. The choice in this case is how long to wait since the last reward, before abandoning each patch. The optimal policy for an animal with perfect timing would be to leave immediately after a FI had lapsed without a delivery, as that was a sure sign that the patch was exhausted, regardless of the mean and variance of travel between patches. However, animals process time intervals with some level of error, well mapped in the field of interval timing. For instance, the standard deviation of birds’ estimates of when food is due, as expressed in the temporal location of the peak in their pecking rate, is known to be proportional to the interval lengths. This is a regularity known in psychophysics as Weber’s Law (see Vasconcelos et al., this volume). This means that as time waiting for a prey item lapses, the bird gets a gradual rather than stepwise increase in its level of certainty about the patch status. This gradual increase in the certainty that the patch is depleted transforms the task into one of choosing between a declining function expressing the potential of the present patch as a function of waiting time and the expected reward rate in the environment. The lower the environmental rate, the longer should the optimal decision maker wait before leaving. This task is equivalent to the problem of when to stop loading prey in the MVT, but for different reasons: instead of the local reward potential declining objectively as a function of time since arrival, here the estimated reward potential declines gradually since the last capture due to the animals’ psychology. The authors reasoned that starlings should approximate the optimal solution taking into account these constraints. To test this, they manipulated inter-prey intervals and travel time between patches. Across conditions they corroborated that the predictions of this psychologically determined version of the MVT were upheld, as follows. First, maximum pecking rate was consistently centered around the FI parameter but the spread of the timing function increased with the length of the FI, as expected from Weber’s Law for timing. This was incorporated as an assumption of the model. Second, the time between the last reward and the time at which birds stopped responding or initiated a new travel kept a roughly linear relation to the FI. This makes sense because of the Weber Law assumption: If timing accuracy were independent of the FI an optimal animal would give up on a patch after waiting a constant time after the last reward, but one whose precision is inversely proportional to the typical FI (as assumed by Weber's Law) needs to vary its waiting time to reach the same level of certainty about a patch's potential for reward. This interlocking between optimality modeling and psychological research is in our view preferable to *ad-hoc* rules-of-thumb or heuristics, because the properties of psychological timing do not need to be seen as dedicated mechanisms to solve a particular experimental situation.

2.2. Optimality and environmental variability

Thus far we have dealt with models that are only sensitive to average parameters. For instance, in Equation 1, the maximized currency is the ratio of average gains to the average sum of travel plus patch times. Equation 1 is thus insensitive to variability in these parameters, but this is not true for how animals behave. Cuthill and colleagues (1990; see also Cuthill, Haccou, & Kacelnik, 1994), for example, found that when starlings experience a mixture of travel times, patch exploitation is affected by the most recently experienced travel time, thus causing variability in patch times. In a different study, Kacelnik and Todd (1992) compared patch residence time in a MVT task between conditions with equal mean travel time but different variance. They found that pigeons (*Columba livia*) decreased patch times with increasing variance. This result can be understood by reference to a mathematical result called Jensen's inequality, which in its simplest form states that if F(x) is a concave function of x and x is variable, then the mean of F(x) is less than or equal to F(mean of x), with the opposite holding for convex functions. Let us consider a bird foraging in an MVT situation but in either of two conditions differing only in the variability of travel times. In the constant condition C_c, the travel time τ is always the same, whereas in the variable condition C_v there are two equiprobable travel times, $\tau+\delta$ and $\tau-\delta$. The average travel time is the same in both conditions, hence the optimal strategy should be the same under constant or variable travel time, because the rates of return are identical:

$$R(t) = \frac{G(t)}{\tau+t} \equiv \frac{G(t)}{\frac{1}{2}[(\tau+\delta)+(\tau-\delta)]+t} \qquad (1b)$$

Thus, and in contrast with the empirical results showing that birds actually do show cycle-to-cycle variability in patch times when travel time is variable, an ideal model animal should leave all patches after the same patch time. This predicted insensitivity is intuitively important to underscore that the MVT predicts decisions taking into account expected (i.e. future) average opportunities rather than paying the travel costs already incurred (sunk costs).

The ideal forager leaves its current patch when it expects, on average, to get a higher payoff elsewhere, given the statistical properties of the environment. If the environment has a random mixture of travel times, then the mean of this mixture defines the rate of gain to be obtained in the future. But the future can only be anticipated by measuring the statistics of the past, and sensible algorithms for using the past to predict the future include some weight for recency, to take into account that conditions may change. Recency effects in turn can cause persistent modifications of patch time.

Even if these fluctuations in patch time are left aside, variation in mean patch time as a function of variance in travel, such as that described by Kacelnik and Todd (1992) can be attributed to a mixture of optimality and mechanistic considerations. Imagine, for instance, a forager that is sufficiently adjusted to the fact that short and long travels are in an unpredictable temporal sequence that it does not vary its patch time between cycles, but modifies patch time gradually according to the discrepancy between the rate of gain it expects and that experienced in each cycle (McNamara & Houston, 1985). In cycles where the preceding travel was short it estimates that the rate of returns is relatively high, and in cycles when preceding travel was long, that it was lower. On average, its subjective estimate is given by Equation 2:

$$R_v(t) = \frac{1}{2}\left(\frac{G(t)}{\tau+\delta+t} + \frac{G(t)}{\tau-\delta+t}\right) \tag{2}$$

that can be simplified to

$$R_v(t) = \frac{G(t)}{\tau+t-\frac{\delta^2}{\tau+t}} \tag{2b}$$

Given that the only difference between Equations 1 and 2b is that Equation 2b has a smaller denominator, it follows that $R_v(t) > R(t)$, and a forager that computes average rate in the future as a cycle-by-cycle running average will perceive the world as being richer. Because perceived lost opportunity while in the patch is greater in a richer environment, the animal would leave all patches sooner, as reported by Kacelnik and Todd (1992). This is another combination of optimality and psychological considerations. Shorter term fluctuations in patch time as reported by Cuthill et al. (1994, 1990) add another layer of mingling between psychology and optimality.

The differences between strategies predicted on the basis of objective rates and strategies that, due to computational mechanisms, lead animals to optimize respect to biased subjective estimates is related to the *fallacy of the averages* which is another idea related to Jensen's inequality, described above. Someone committing this fallacy fails to see the distinction between computing a mean of a variable and then applying a function to it, and applying a function to each case of the independent variable and then computing the average. (see discussion in Templeton & Lawlor, 1981; Turelli, Gillespie, & Schoener, 1982). In the example above, Equation 1b (a function of expected values) leads to predictions different from those of Equation 2 (the expected value of a function). The original fallacy is only of historical interest, because the issues are now well understood, but the problem of which algorithm predicts behavior remains, and boils down to whether including psychological mechanisms is a fair practice in optimality models. Regarding rate maximizing, the experienced rate of return of an animal over a certain period is given by the quotient between mean gains and mean times over the period, but it is perfectly possible that real foragers instead respond psychologically to the mean of the ratio of gain over time across hunting episodes, thus "committing the fallacy", as it were. Including such mechanisms has methodological costs, because one builds into the model what should ideally be the model's output, but it is to some extent unavoidable if the models are conceived as being refined with relation to behavioral data. In their original conception, optimality modelers assumed that the constituent hypotheses of each of their preferred models were objectively known. The role of experimentation at the time was aimed, in practice, to illustrate the action of natural selection, and to generate predictions for previously unknown behavioral adaptations. The hypotheses involved in constructing the models were not explicitly seen as such, leading to the illusion that functional behavioral models could ignore mechanisms.

3. The self-control problem in intertemporal choice

Another research field that benefits from integrating functional and mechanistic approaches is the study of self-control in intertemporal choice. Intertemporal choices involve decisions between outcomes at different times in the future, and they are pervasive in human and non-human daily life. Consider once again a foraging bird that provisions its chicks, this time facing a choice between two feeding patches, one providing a large prey after a long search time (LL) and the other a small prey after a brief search (SS)[3]. For simplicity let's assume that the bird carries only one prey per trip. Since the bird spends all its daytime provisioning its young, natural selection favors individuals that maximize the food mass

---

[3] The reader may notice some similarities between this situation and the patch-exploitation problem (for a detailed analysis see, for example, Stephens & Anderson, 2001; Stevens & Stephens, 2010).

delivered over the day. We can now compare the adaptive advantages of choosing to hunt for LL or SS. Since the birds are judged by the amount of food delivered, they would be indifferent when the ratio of reward size to search time plus travel time is the same in both sites:

$$\frac{Size_S}{\tau+t_S} = \frac{Size_L}{\tau+t_L} \tag{3}$$

or

$$Size_S = \frac{\tau+t_S}{\tau+t_L} Size_L \tag{3b}$$

In other words, the biological value of a large reward is equivalent to that of a small reward provided that the ratio of times involved exactly compensate for the size difference. One way to express this is to say that delayed rewards are discounted as a function of the extra time costs. Equations 3 and 3b are based on the assumption that time is a limiting resource.

Many laboratory studies examine intertemporal choice in animals using protocols wherein subjects, typically pigeons or rats (*Rattus norvegicus*), choose repeatedly between alternatives that differ in reward size and delay in trials separated by intertrial intervals. In a prototypical example, as soon as a trial starts a pigeon chooses between two colored keys, each associated with a particular amount and delay to food. SS may give one pellet of food after a delay of 10 s and LL two pellets after 20 s. In different conditions, experimenters manipulate the delays to and/or amounts of food so as to map how animals trade amount for time. Given these values, if there is no ITI (or if the subjects do not include it in their computations, as we discuss later on), according to Equation 3b the two items should be equally valuable, but this has not been the intuitive expectation in most of the psychological literature on temporal discounting.

In treatments of this problem by experimental psychologists, although in their experiments the animals face iterated choices, the two most frequent mathematical descriptions consider the choices as if the animal made only one choice. The question asked is not “which relation between size and delay would equalize rate of gain?” as is typical in optimal foraging theory, but instead “what function describes the value of a reward as a function of waiting time?” This function is normally called the discounting function. Because of the implicit one-shot perspective, lost opportunity is not in the frame, and the results are often described saying that animals forego long-term gains in favor of more immediate but less valuable food rewards (e.g., L. Green, Fisher, Perlow, & Sherman, 1981; Mazur, 1987;

McDiarmid & Rilling, 1965), as if this were irrationally impulsive (the phrase “inability to delay reinforcement” is sometimes used). Pigeons, for example, are said to find rewards delayed just one second as half as attractive as an immediate reward (Mazur, 1984). According to the rate of reward analysis, if it were possible (and this is of course not the case) for a reward to be found and consumed in no time at all, the corresponding rate of gain would tend to infinity, and there is no size of a more delayed reward that equalizes the value of an immediate delivery.

Within the one-shot framework there are both normative and descriptive accounts of temporal discounting. Normative (i.e. optimality) models of temporal discounting, suggested both by economists (Samuelson, 1937) and biologists (Kagel, Green, & Caraco, 1986), predict that the discounting function should take an exponential form. This is because if there is a constant probability of the reward being lost per unit of waiting time, the probability of collecting the reward is a declining exponential function of the delay. This is known as the “discounting-by-interruptions” hypothesis, and has been central to treatments of intertemporal choice in behavioral ecology (e.g., Houston & McNamara, 1999; Kagel, Battalio, & Green, 1995; Kagel et al., 1986; Sozou, 1998). This is perhaps surprising as the one-shot logic contrasts with the classical optimal foraging treatments that see time as a limiting resource and focus on multiple decisions and repeated cycles.

A descriptively successful and widely accepted alternative to exponential discounting, termed hyperbolic discounting, has been put forward by Mazur (1987). According to this model, value declines with delay as follows:

$$V_i = \frac{S_i}{1+kt_i} \tag{4}$$

where S_i is the subjective value of the reward if it were available immediately, k is a free parameter with dimensions reciprocal to time and t_i is the delay between the evaluation moment and the outcome for that prey. Variations in k are used to describe differences between individuals and between species. This function has been successful in fitting animal choice data and useful in clinical settings, where research has shown an association between the parameter k and addictive behavior, such as gambling, substance abuse, and obesity (Odum, 2011; Odum & Baumann, 2003).

Although derived descriptively rather than normatively, hyperbolic discounting is almost identical to Equation 3. Let us assume that the subjective value of a prey immediately available is well represented by its size. Then the delays at which two prey LL and SS have equal value are given by

$$\frac{S_S}{1+kt_S} = \frac{S_L}{1+kt_L} \tag{5}$$

or

$$S_S = \frac{1+kt_S}{1+kt_L} S_L \tag{5b}$$

In addition to the fact that Equation 3b has a normative interpretation under repeated cycles while 5b is descriptive and frames discounting as a one-shot problem, the main differences between them are as follows. In Equation 3b the proportionality constant is non-dimensional because it is the ratio of two times, while in Equation 5b the time dimension is eliminated by the fitted constant k. Equation 3b includes the ITI, while Equation 5b replaces it by a small constant that is influential only for very short delays. And finally, 5b has a fitted parameter that allows for different discounting rates across subjects or species, while Equation 3b is rigid in this respect.

There remains a further consideration that will serve us, once again, to examine the relation between optimality models and what we may call “real” psychology. Equation 3b implies equal sensitivity to all time components, namely the delay between choice and outcome *t* and the travel time or inter-trial interval τ, but, because they deal with one-shot problems, Equation 5b only addresses sensitivity to the delay between choice and outcome. Available evidence, although scarce, suggests that intertrial intervals have very little effect in animal self-control experiments (Mazur & Romano, 1992) or other designs including choices between simultaneous opportunities. This is puzzling, because in the patch-exploitation problem, travel time, which can be seen as closest to the intertrial interval in self-control studies, has a strong and highly predictable impact. Consistently with the view expressed elsewhere (Kacelnik, 2003) we argue that the answer lies in the temporal position of the time components of the cycle relative to the moment at which the subject makes its choice. In the patch-exploitation, central place foraging problem used to describe the MVT, the forager’s decision is when to leave the current patch to initiate a cycle by traveling to the nest, hence travel costs occur between the decision and its consequences, the reinforcing experience of arrival at the nest and then at a new foraging site. In the self-control paradigm, the decision is the choice between SS and LL and the delays occur between that decision and its outcome, with the ITI placed after the choice consequence (Figure 3 illustrates our reasoning).

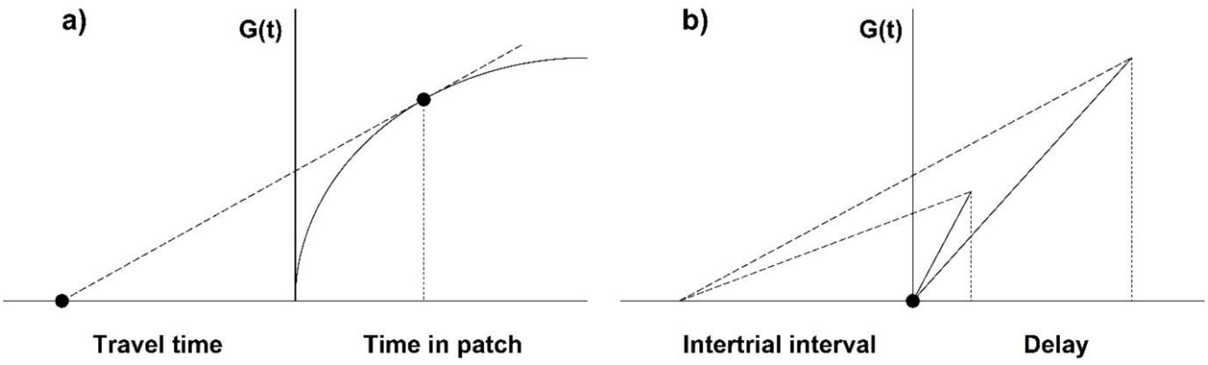


Figure 3. The effect of travel time and intertrial-interval (ITI) on decisions. (a) In the depleting patch problem, the decision being modeled is patch departure, which marks the start of a new foraging cycle. Travel time occurs between the decision, indicated by the two black dots, and a new patch arrival. Given the cyclical nature of the problem, the two dots mark the same point in time; (b) in the discounting problem, the decision between two options and the programmed delay occurs between the decision, occurring at the moment indicated by the black dot, and either outcome. Travel time (or ITI in lab simulations) occurs between outcomes and new choices. Although the overall rate of reward, as indicated by the slope of the broken lines, shows that travel time has the same effect on rate of reward in both the depleting patch and the discounting problems, reinforcement analysis expects them to be substantially different, and predicts that travel time will control decisions in the MVT while ITI will be irrelevant in the discounting problem.

Our view is that animals are very sensitive to times between decisions and outcomes, but relatively insensitive to intervals other than those, because of the problem of *credit attribution*. From an evolutionary standpoint, it makes sense that when the animal obtains a reward (SS or LL) after a short or long delay respectively it attributes the “responsibility” of such outcome to its preceding decision. It may be expected, both on conditioning and foraging arguments that the reinforcing effect of the outcome declines with the interval since that decision, and increases with the magnitude of the outcome. This is adaptive in a world where the animal needs to learn the consequences of its decisions by the ensuing outcomes (e.g., Bouton, 2007).

With suitable adjustments, Equation 3 can be applied to situations where the effect of energetic costs on choice is examined. Consider a situation where the forager faces alternatives that differ in net energy content S_i (net content is the absolute content minus the metabolic expenditure during handling time), handling time t_i, search time τ_i, and metabolic rate during searching, m_i. In such a case, the net rate that would result from using exclusively option i is given by:

$$NetR_i = \frac{S_i - m_i\tau_i}{\tau_i + t_i} \qquad (6)$$

This model was applied by Bautista et al. (2001) in a laboratory study of starlings choosing between searching for food by flying or by walking. The question was, given that flying is more metabolically expensive than walking, how would the birds integrate time and metabolic costs to make their choices, starting with the prediction that they might be expected to maximize Equation 6. The birds were given iterated choices between two options delivering food rewards of equal size, one requiring a certain (manipulated) length of time flying and the other an adjustable length of time walking. For each experimentally fixed flight time, the walking cost was automatically varied to establish by titration the value at which the birds were indifferent between the options, considering that flying delivered food faster but at greater cost per unit of time than walking. The results were also compared to two alternative currencies, gross rate of energy gain (ignoring metabolic costs) and energetic efficiency (energy gained per unit of energy spent, ignoring the times involved). The results were very close to those predicted by Equation 6, indicating that the birds do indeed include both time and energy costs in determining their preferences between sources of food. This is another form of hyperbolic discounting, because by flying a bird gets expensive food sooner while by walking it gets cheap food later, but the protocol is enriched by the inclusion of energetic costs to reflect the foraging perspective.

In summary, hyperbolic discounting is the form of discounting in intertemporal choices predicted by optimality models based on rate maximization. Some deviations from its predictions in special cases are as expected by learning processes, where decisions are reinforced by their consequences. With one exception, to our knowledge, single-shot choices between SS and LL rewards are not appropriate to model animal choices, given that the animals are instructed of the parameters of the protocols by iterating multiple trials. The exception is Stevens, Rosati, Ross, and Hauser (2005) work with cotton-top tamarins (*Saguinus oedipus*) and common marmosets (*Callithrix jacchus*), where delay could be anticipated from visual cues. More generally, many results of choice experiments with non-humans can be explained by identifying the decision facing the subject and considering the way actions are reinforced by their consequences. The overwhelming selection pressures may have been those designing the existing reinforcement mechanisms, rather than specific solutions to unique choice protocols, and for this reason the search for rules of thumb may not be the most fertile approach to relate optimality models to psychological processes.

4. The structure of foraging environments and choice

A common assumption in the study of decision making is that most, perhaps all decisions imply tradeoffs. As we have just discussed, animals need to choose between foraging opportunities that differ in magnitude and time costs. From such a starting point it is tempting to assume that decisions involve comparisons between alternatives and therefore cognitive effort and time. As we will see, this assumption is not always supported.

We incorporate the hypothetical structure of foraging environments in our modeling approach by arguing that animals mostly choose in contexts where alternatives are faced sequentially rather than simultaneously. An example is when a predator decides between pursuing a detected prey or skip it and keep searching for alternatives that may be more profitable because they are larger, less likely to escape, or require less engagement time and hence less lost opportunity. The ideas in this section are encapsulated in the Sequential Choice Model (SCM; Kacelnik, Vasconcelos, Monteiro, & Aw, 2011; Shapiro, Siller, & Kacelnik, 2008) whose main assumption is that choice mechanisms evolved as adaptations to environments in which sequential encounters (i.e., finding one option at a time) are common, whilst direct choice opportunities (i.e., finding two or more options simultaneously) are rare (e.g., Stephens & Krebs, 1986). Thus, the psychological mechanisms acting in choices between simultaneously available options are adaptations for the broader need of learning to improve performance in more frequent situations, such as when deciding whether to engage with a given option or let it pass to pursue other alternatives (closer to a go-no go psychological protocol).

According to this framework, animals develop a subjective valuation of each source of reward (similar to its associative strength in learning models) whenever they encounter that prey type, through the mechanisms of reinforcement learning. With the same lost opportunity reasoning used so far, this valuation is a function of the remembered profitability of each prey type relative to the rate of gain in the environment as a whole, including time costs and information about the mixture of alternatives in the habitat. We further hypothesize that when animals face single options, relative valuation is expressed as the immediacy of their response ("latency"). In nature this latency would correlate with the probability of skipping the opportunity to search for alternatives. Latencies in encounters with single options should decrease when the options' objective profitability increases (e.g., ratio of amount of reward to the delay between action and outcome; Bateson & Kacelnik, 1996; Mazur, 2010; Shapiro et al., 2008), and increase with both the energetic reserves of the animal during learning (e.g., Aw, Holbrook, Burt de Perera, & Kacelnik, 2009; Pompilio & Kacelnik, 2005; Pompilio, Kacelnik, & Behmer, 2006; Vasconcelos & Urcuioli, 2008) and the profitability of the available alternatives in the same context (Fantino & Abarca, 1985; Mazur, 2010; Shapiro et al., 2008). Due to random noise, successive sequential encounters with an option result in a probability density function of latencies. We further hypothesize that when two or more options are met simultaneously, each option elicits a sample from its own distribution of latencies and the shortest sample is expressed as a choice.

In this hypothetical mechanism no deliberative comparison between options occurs at the time of choice; instead, the mechanism underlying sequential decisions is primed for each option in parallel, with the option yielding the shortest sample dictating behavioral allocation and censoring the alternative (Kacelnik et al., 2011). Formally, the probability P_A of choosing option A over B is given by the joint probability of the latency for A equaling *x* and the latency for B exceeding x, integrated for all possible *x*:

$$P_A = p(l_A < l_B) = \int_0^{\infty} f_A(x).\,[1 - F_B(x)]dx \qquad (7)$$

where l_A and l_B are random samples from the respective distributions, f_A is the probability density function of latencies for A, F_B is the cumulative distribution function of latencies for B, and *x* is a particular latency value.

In summary, the SCM's assumptions are that the latency to accept sequentially encountered options is a joint function of three variables: the options' objective properties, the average rate of gain in the environment given the mixture of options and their respective search times, and the energetic state of the subject at the time of learning (i.e., not at the time of the choice). Moreover, the mechanism that determines choice when two or more options are met simultaneously depends on random independent sampling from each option's latency distribution. Under this hypothetical mechanism the latency distribution of the options present in a simultaneous choice censor each other, as only the shortest of the sampled latencies will be recorded. In other words, the SCM is what is known as a race model in the decision making literature, as opposed to models postulating a competitive evaluation between attractors such as diffusion models (Bogacz, 2007; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Pelé & Sueur, 2013).

The assumptions of this model combine empirical observations and the overarching logic of optimal foraging theory. One departure from *a priori* optimality that is empirically driven is the very existence of a latency to respond to single options. If (as it happens in many laboratory experiments) a predator has no choice other than take the present option or remain in that situation forever, it should take it immediately. Typically consuming the prey immediately resets the scene to searching anew, meaning that any waiting time is a loss of opportunity. In spite of these expectations, Shapiro et al. (2008; see also Mazur 2010) found that latencies in starlings facing exactly those circumstances were strongly dependent on the parameters, namely were shorter when the present option was richer and when the alternative option in the environment was leaner (Figure 4). Post-hoc arguments for the adaptive significance of these latencies have been elaborated since, for instance the suggestion that latencies in the laboratory are an artifact of the animals' lack of choice: in nature foragers can always skip a prey to forage for further opportunities, and the conditions under which a prey should be skipped are exactly those that cause longer latencies in single encounters in the laboratory (Charnov, 1976a). Whatever the adaptive history of the mechanisms that cause such orderly latencies, once their presence is established optimality models need to include them in the strategy set, lest the models are condemned to failing from scratch. Models including these assumptions can make novel and counterintuitive predictions. For instance, according to this model it ought to be possible to predict preferences in simultaneous choices from the distributions of latencies in sequential encounters.

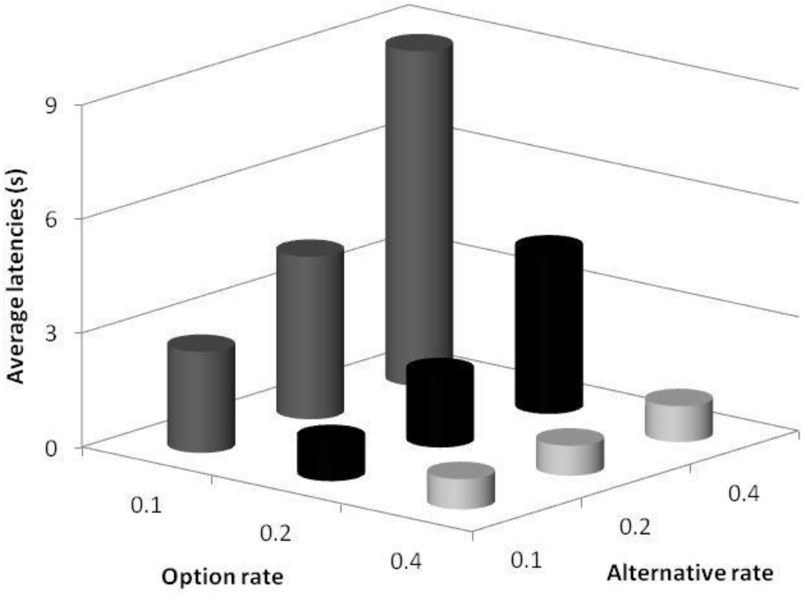


Figure 4. Median latencies to accept each option during sequential encounters as a function of the option's profitability and that of the alternative option that could potentially be encountered in the same environment, averaged across subjects. Adapted from Shapiro et al. (2008).

The SCM predicts a deviation from indifference in simultaneous choices if and only if different latency distributions are observed during sequential encounters. The option with the most leftwards distribution of latencies in sequential encounters should be chosen more often, and thus be less severely censored than the alternative. The less preferred alternative will seldom be chosen, and when it does the observed latencies should on average be shorter than when the same option is encountered alone. Through this censorship mechanism the SCM makes the counterintuitive prediction that latencies observed in simultaneous choices should be shorter than the ones observed with those same options in sequential encounters. Moreover, the shortening of latencies in choices with respect to sequential decisions should be more extreme for the less relatively profitable and less frequently chosen option.

The prediction of a latency shortening contradicts models of choice that assume the existence of comparative cognitive processes at the time of making each choice. Such comparisons should logically take time, and thus animals would be expected to act faster when options are met without competition than when multiple options are presented simultaneously. This intuitively expected increase in choice time with number of options is encapsulated in the Hick-Hyman Law (Hick, 1952; Hyman, 1953). The assumption is that every choice involves a trade-off between accuracy (i.e., choosing the better option) and the (temporal) cost of evaluation.

Similarly to other proposals for individual decision making (e.g., Blough, 2011; Ratcliff, Van Zandt, & McKoon, 1999; Smith & Ratcliff, 2004) and collective decision making in social insects (Robinson, Franks, Ellis, Okuda, & Marshall, 2011; Seeley et al., 2012), the SCM proposes that choice processes resemble a horse race rather than a tug-of-war. The idea is that the vertebrate's brain operates at least partly like a eusocial insect colony, where options are represented by some form of bidding in the absence of an executive system that ponders their relative strength.

The model was inspired by experimental results obtained with starlings foraging in two-alternative environments (Shapiro et al., 2008), but has since been successfully applied to risk-sensitive foraging (Aw, Monteiro, Vasconcelos, & Kacelnik, 2012) and multi-alternative environments (Freidin, Aw, & Kacelnik, 2009; Vasconcelos, Monteiro, Aw, & Kacelnik, 2010). Simultaneous choice can be predicted in multiple ways under the SCM rationale. One, which we call molar, uses the complete distribution of latencies from sequential encounters with each option to predict the overall proportion of choices. The alternative, which we call molecular, attempts to predict the outcome of each particular choice using only the most recent sequential encounters with each option, thus considering potential local fluctuations in motivation. Figure 5 shows experimental results and model predictions in two- and multi-alternative environments. Figure 5a includes only molecular predictions while Figure 5b includes both molecular (left axis) and molar predictions (right axis).

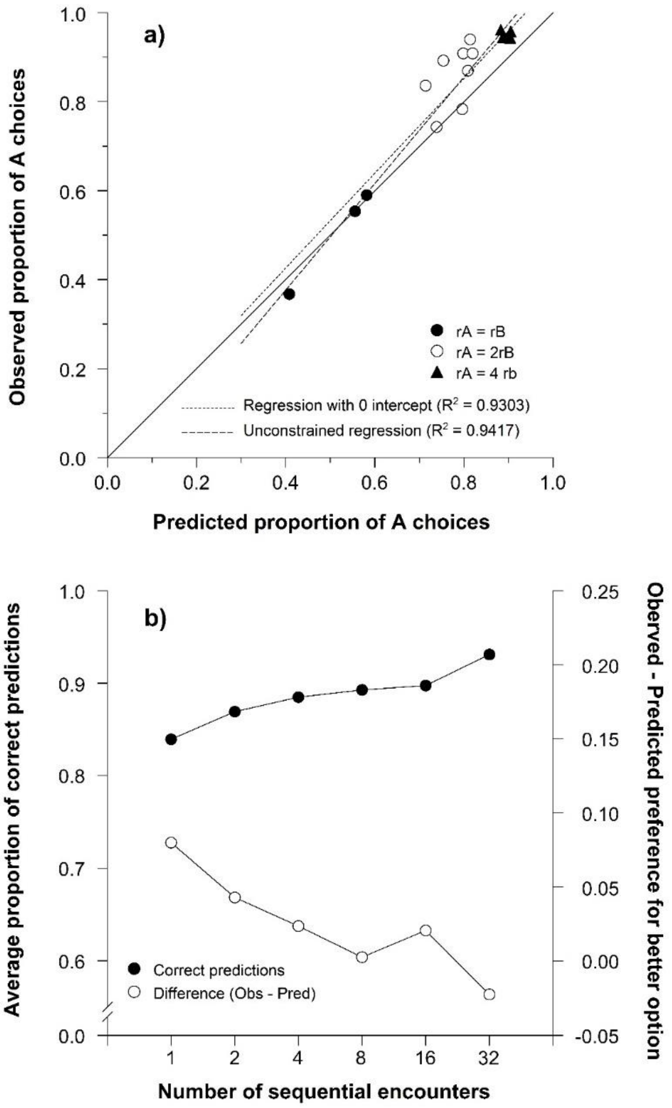


Figure 5. Illustration of the SCM ability to predict simultaneous choice from sequential encounters. (a) Obtained versus predicted proportion of choices for the option yielding higher rate of reward (always labelled A) according to the SCM. Each rate is represented by a different symbol. Two linear regression lines are included. The dashed line corresponds to an unconstrained regression and the doted one is constrained to pass through the origin (adapted from Shapiro et al., 2008); (b) Average proportion of choices accurately predicted by the SCM (solid circles, left axis, molecular predictions), and difference between observed and predicted preference strength (open circles, right axis, molar

predictions) as a function of the number of preceding sequential trials used to predict each choice in simultaneous presentations (adapted from Vasconcelos et al., 2010).

Support for the prediction that latencies for each option should be shorter in the presence of alternatives than when the option is alone so far is weaker, but two features make testing this prediction difficult. First, latencies have a lower but not an upper limit, which frequently constrains the detection of shortening by floor effects, unlike the opposing lengthening hypothesis which is not limited by physical constraints. Second, the best chance to observe shortening is in the less preferred option because it is more severely censored, but by definition this option is chosen infrequently and thus the sample size of latencies for this option in simultaneous choices is typically small. Nonetheless, shortening (Shapiro et al., 2008) or a tendency in that direction (Mazur, 2010; Vasconcelos, Monteiro, & Kacelnik, 2013) has been observed on several occasions, while signs for the increase in choice time expected from the hypothesis that choice takes some deliberation time have not been reported so far.

Summing up, the basic idea of the SCM is that choices made when facing alternatives simultaneously can be predicted by behavior observed during sequential encounters with one alternative at a time, but not the other way around. The significance of sequential encounters and the logic of lost opportunity is an ubiquitous feature of optimality models since the early days of optimal foraging theory, through Charnov’s “Diet Choice” model (Charnov, 1976a) and “Marginal Value Theorem” (Charnov, 1976b), both of which predict preferences as a consequence of a tradeoff between exploiting the current source of reward and the background average opportunities offered by the environment as a whole; taking the present option causes lost opportunity elsewhere. The main contrast with psychological accounts of choice is that the latter must be applied to the individual’s experience, and the MVT explains behavior as if it were shaped by the trans-generational species’ past so as to maximize reward in the future. It should be apparent that it makes no sense to explain behavior resorting to only one or the other. Charnov’s ideas from early on inspired empirical tests (Cowie, 1977; Krebs, Erichsen, Webber, & Charnov, 1977) that are now classic in behavioral ecology, but these early tests and most of those that followed did not, in our view, make sufficient contact with psychological research. In experimental psychology, meanwhile, the study of decision making has been dominated by descriptive models of simultaneous choice paradigms, without sufficient interest in the adaptive consequences of animal preferences. Re-focusing on the contrast between single options and their background context, and incorporating known psychological mechanisms into the strategy set of optimality models allows for a productive link between these diverse approaches.

5. Conclusion

We have argued that because behavior and cognitive processes are products of evolution by natural selection, behavioral research can benefit from optimality modeling. This approach assumes that the psychology of organisms makes sense from a fitness-maximization design standpoint, but does not amount to say that animals always behave optimally. Instead the models are sets of hypotheses subject to empirical refinement.

We have illustrated these ideas with optimality models of foraging behavior, one dealing with the decision on when to switch from a reward source with diminishing returns, another with intertemporal choice and discounting, and finally one considering how choices between simultaneously presented opportunities reflect mechanisms evolved to more ecologically relevant sequential choices.

It should be clear that in our view, while optimality modeling uses an evolutionary logic, at its best it is supported by specific psychological hypotheses that are directly tested experimentally. For example, the MVT typically is used to discuss the functional problem of foraging strategies for gradually depleting patches, but because each testable implementation requires hypotheses for a strategy set, a feedback function and a currency, it cannot detach itself from behavioral mechanisms. These components are hypotheses that can be rejected empirically, and in proper implementations they often are. Data can show that the feedback function is inaccurate or that the animal maximizes a different currency, and combined laboratory experiments with field work are needed to disentangle these possibilities. Models respond to data by being modified to include previously unforeseen constraints, changing the currency, or specifying the statistical structure of the foraging environment more precisely, and making new, testable predictions on these basis. Good models account for what is known and predict what hasn’t been explored as yet. What persists is the models’ functional rationale: that the biological system, including its psychological mechanisms, reflects the action of natural selection.

It is widely appreciated that testing environments seldom match precisely the environment under which the behavioral mechanisms evolved (the problem of the domain of testing vs. the domain of selection; see, for example, Stevens & Stephens, 2010). This can be tackled using artificial selection experiments. For instance, Dunlap and Stephens (2009) have succeeded in using optimality considerations to anticipate the rapid evolution of learning parameters across just 30 generations of fruit flies (*Drosophila melanogaster*). But even when optimality predictions tested against current ecological conditions may fail because of environmental mismatching, the models provide a strong framework for the study of behavioral mechanisms. Functionally inspired models help understanding animals' responding to experimental protocols, by relating them to possible natural equivalents, for instance relating intertrial intervals to travel time, food caching to spatial memory tasks, perception of probability and partial reinforcement to risk sensitivity, and so on.

In our view, a truly integrative study of animal behavior and cognition must combine evolutionary and psychological approaches. They are often seen as orthogonal topics, but through model-inspired experimentation they can and do converge.

CHAPTER III[4]

STUDY 1:

IGNORING "BAD NEWS": PIGEONS DISREGARD STIMULI THAT ARE NEVER FOLLOWED BY FOOD

---

[4] This Chapter reproduces the submitted version of the paper:

Fortes, I., Vasconcelos, M., & Machado, A. (2016). Ignoring "Bad News": Pigeons Disregard Stimuli That Are Never Followed By Food. *Manuscript submitted for publication.*

Abstract

Several studies have shown that pigeons prefer options followed by informative stimuli indicating the presence or absence of food, rather than options followed by non-informative stimuli concerning the impending delivery of food. In this paper we tested Vasconcelos, Monteiro and Kacelnik's (2015) model that predicts that when an option includes a stimulus that signals the absence of food, the stimulus is ignored in the sense that its probability and duration have no effect on the value of the option. To that end, pigeons chose between two options: the Informative Option delivered food on 20% of the trials after a 10-s delay, signaled by a red key ("good news"), and delivered no food on the remaining 80% of the trials, signaled by a green key ("bad news"). The Non-informative Option delivered food after 10 s on 50% of the trials, regardless of the signal shown (yellow or blue). In Experiment 1, the probability of the "bad-news" stimulus was manipulated from 0.80 to 1.00; in Experiment 2, the duration of the "bad-news" stimulus was increased every time pigeons preferred the Informative Option. Consistent with the model, the results showed that both manipulations had little or no effect on choice.

Keywords: choice, suboptimal, bad news, conditioned reinforcement, pigeons

1. Introduction

Research on choice has shown that animals usually prefer situations in which reinforcers are signaled over those in which they are not; that is, they prefer to know in advance if a reinforcer is due or not (e.g., Bower, McLean, & Meacham, 1966; Prokasy, 1956; Roper & Zentall, 1999). For instance, Prokasy (1956) found that, in an E-shaped maze, rats preferred an arm that was white when it was baited and black when it was not, over an arm that was white or black regardless of the presence of food, even though the probability of food was the same in the two arms. In a similar experiment (Roper & Zentall, 1999), pigeons chose between two keys associated with the same probability of reinforcement. If pigeons chose the Informative Option, on 50% of the trials they saw a positive discriminative stimulus (a "good-news" stimulus) that was always followed by food after 6 s; on the other 50% of the trials they saw a negative discriminative stimulus (a "bad-news" stimulus) that was presented for 6 s but always ended without food. If they chose the other option, the Non-informative Option, they saw one of two stimuli, both uncorrelated with food. That is, after a 6-s delay, food was delivered on 50% of the trials regardless of the stimulus presented. Roper and Zentall found a strong preference (above 90%) for the Informative Option.

In these experiments, the animals preferred the option that informed them about the forthcoming outcome but their preference had no cost because the overall probability of reinforcement was the same in both options. Subsequent research has shown that, under certain boundary conditions, at least two species (pigeons, *Columba livia*, and European starlings, *Sturnus vulgaris*) "trade" food for information. For example, Stagner and Zentall (2010) found that even when the good-news stimulus was presented on only 20% of the trials, that is, when the overall probability of food in the Informative Option was 2.5 times lower than in the Non-informative Option (20% vs. 50%, respectively, see Figure 1), pigeons strongly preferred the Informative Option. Many other studies have reported analogous tradeoffs, including Kendall (1974), Mazur (1995, 1996), Spetch, Mondloch, Belke and Dunn (1994), Stagner, Laude and Zentall (2012), Stagner and Zentall (2010) and Vasconcelos, Monteiro and Kacelnik (2015) to name just a few (for a review see, Zentall, 2014, 2016).

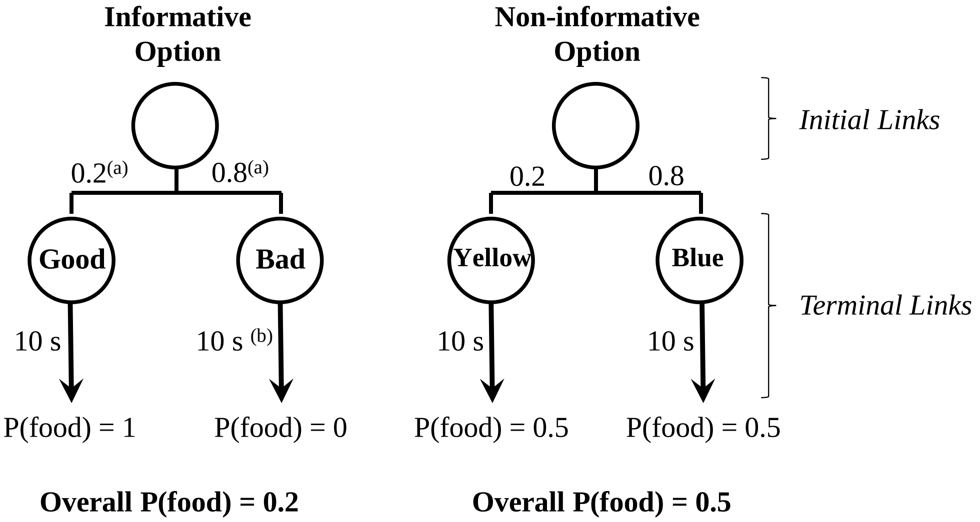


Figure 1. Schematic of outcomes for each option in Experiments 1 and 2. In Experiment 1, the probabilities (a) were manipulated. The probability of bad news was 0.80, 0.90, 0.95, 0.975 and 1.00. In Experiment 2 the delay (b) was increased when pigeons preferred the Informative Option and decreased when they preferred the Non-informative Option. The probability of bad news was kept constant at 0.80.

Why would animals choose an option yielding less food than the alternative? The seemingly obvious explanation is that animals avoid the Non-informative Option because it is ambiguous, whereas the Informative Option provides immediate information about the delivery of food. To test this ambiguity avoidance hypothesis, Zentall and Stagner (2011) removed the ambiguity from the Non-informative Option by making it deliver the same amount of food on *every* trial (3 pellets); the Informative Option delivered a larger amount of food (10 pellets) but on only 20% of the trials. Even though the Non-informative Option was now unambiguous and yielded more food on average (3 vs. 2 pellets), pigeons still preferred the Informative Option. Thus, they do not avoid the optimal alternative because it is ambiguous.

Another hypothesis is not centered on the aversiveness of the ambiguity associated with the Non-informative Option, but rather on the attractiveness of the conditioned reinforcer provided by the Informative Option. In other words, the stimuli associated with the Non-informative Option are weak conditioned reinforcers because they are only moderately correlated with food. On the other option, the good-news stimulus may act as an enhanced conditioned reinforcer because it is perfectly correlated with food. On the other hand, on the same option, the bad-news stimulus, which is presented much more frequently than the good-news stimulus - usually 80% of the time - leads to a high proportion of unreinforced Informative trials, which should turn it into a conditioned inhibitor. As a consequence, if the good-news stimulus adds value to the Informative Option, the bad-news stimulus should reduce the value of this option. However, there is evidence that this is not the case because the bad-news stimulus does not have inhibitory effects. Laude, Stagner and Zentall (2014), for example, showed that even though the bad-news stimulus may start as an inhibitory stimulus, it loses the inhibitory properties with training.

Given that the bad-news stimulus does not seem to have inhibitory or punishing properties at steady state, it has been claimed that a negative discriminative stimulus plays little role in choice (Mazur, 1991, 1995, 1996; McDevitt, Spetch, & Dunn, 1997; Spetch et al., 1994; Stagner et al., 2012; Zentall, Laude, Stagner, & Smith, 2015). If that is the case, in the choice task depicted in Figure 1, pigeons choose suboptimally because they are functionally choosing 100% reinforcement (in the Informative Option, disregarding the bad-news stimulus trials) over 50% reinforcement (in the Non-informative Option).

If animals do ignore the bad-news stimulus, manipulations of its *probability* and *duration* should not affect preference for the Informative Option. This prediction may seem paradoxical, because as the probability or duration of the bad-news stimulus increases, the rate of food intake on the Informative Option decreases, so optimal animals should prefer this option less. On the other hand, if the bad-news stimulus is not taken into account, its associated probability and delay are irrelevant for the perceived rate of reinforcement, so animals' preferences should not be affected. In the present paper we aimed to test these predictions and explore their boundary conditions.

Concerning the probability of bad news (and, as a consequence, the probability of good news), there is evidence suggesting that it has no effect on choice (Mazur, 1996; Stagner et al., 2012; Vasconcelos et al., 2015; Zentall et al., 2015). Zentall et al. (2015) showed that preference was equally suboptimal when the probability of good news was 25% or 50% while the other alternative provided unsignaled food on 75% of the trials. Vasconcelos and colleagues (2015) reported even more noteworthy evidence: When they progressively decreased the probability of good news from 20% to 0% in steps of 5%, they found that starlings preferred the suboptimal, Informative Option reliably above chance even when it yielded food on only 10% of the trials, while the Non-informative Option provided food on 50% of the trials. In this case starlings lost 80% of the potential rewards. In a nutshell, animals seem to ignore the probability of bad news in the Informative Option, taking into account only the trials in which food delivery was signaled (the good-news trials).

Based on an optimal foraging analysis (Stephens & Krebs, 1986), Vasconcelos et al. (2015) proposed that in their natural environment, when animals contact with cues signaling the absence of food, they do not pay the cost of waiting; instead, they use the information that no food is available to search for other opportunities. However, this strategy will backfire in any laboratory task that forces the animal to wait in the presence of a signal for the absence of food: In experimental tasks such as that depicted in Figure 1, animals cannot use the information imparted by the bad-news stimulus by, for example, moving away to search for another source of food, thus ending up paying the cost of waiting. Because animals may apply the same strategy within and without the laboratory – ignore the bad-news stimulus – the cost of waiting in the presence of the bad-news stimulus in the laboratory is ignored and is not associated with the chosen option. In these circumstances, the behavior of the animals will indeed be suboptimal, because they are using a mechanism which evolved and it is advantageous in their natural environment but has no advantage in the laboratory.

Consider an animal that maximizes the rate of food intake. While foraging, the animal starts by *searching* for a food item. When an item is found, the animal stops searching and starts *pursuing* the item. After some time, the animal may catch the item and start consuming it (a *handling* time) or it may lose the item, thus ending the foraging cycle and starting a new one by searching again. If the animal includes all the delays in the rate computation, the rate of return (R_i) that it gets if it chooses exclusively one food source or prey type is

$$R_i = \frac{p_i}{s + p_i \times (t+h) + (1-p_i) \times t} \tag{1}$$

where R_i is the rate of return in energy/time from option i, p_i is the probability of receiving food or capturing the prey, s is the searching time, t is the pursuing time and h is the handling time. For simplicity and to keep the analogy with the task shown in Figure 1, Vasconcelos et al. (2015) assumed that (a) s corresponds to the intertrial interval (ITI), (b) t is the delay associated to each terminal link (cf. Figure 1), (c) h is the time to consume the reward, and (d) t and h are equal across options. Because R_i monotonically increases with p_i, the values typically used in the task ($p_{Info} = .2$ and $p_{Non\text{-}info} = .5$) yield $R_{Non\text{-}info} > R_{Info}$, and thus the animal should prefer the Non-informative Option. In fact, Equation 1 predicts that animals should always prefer the option yielding the higher probability of reinforcement. Figure 2a shows how preference should change as both p_{Info} and $p_{Non\text{-}info}$ vary, with t and s

held at 10 s (the value of *h* has no effect on preference). The white circle corresponds to the typical probabilities used (e.g., Stagner & Zentall, 2010, Zentall & Stagner, 2011, Stagner et al., 2012) and the black circles correspond to other probability combinations tested by Vasconcelos and collaborators. In this case, the predicted and the obtained preference are at odds: These rate functions predict preference for the Non-informative Option in all conditions, but animals always preferred the Informative Option, except when $p_{Info} = 0$.

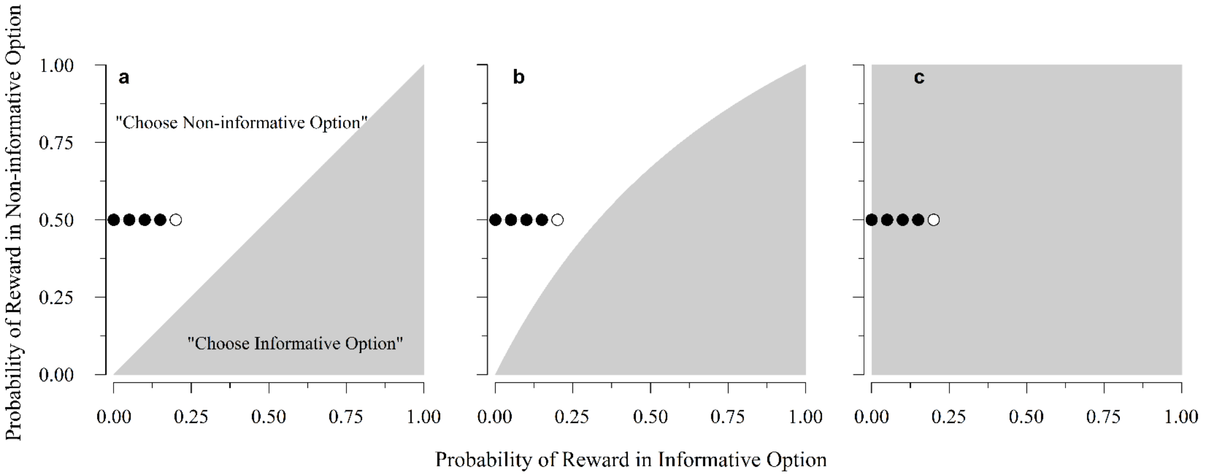


Figure 2. Contour plot of the preferences predicted by the Vasconcelos et al.'s (2015) model in the entire parameter space of $p_{\mathrm{Info}} \times p_{\mathrm{Non\text{-}info}}$. The shaded and white areas indicate preference for the Informative and Non-informative Options, respectively. In all panels, t = 10 s, the black and white circles represent conditions where $p_{\mathrm{Non\text{-}info}} = .5$ and p_{Info} varied in steps of .05, from .00 to .20. (a) Predictions using rate functions including all delays; (b) Predictions using rate functions without delays leading to no reward; (c) Predictions using rate functions without delays leading to no reward and without ITIs.

Consider now an animal sculpted by natural selection to use information and therefore search for alternatives when sure of no impending reward. Such an animal would ignore the time spent waiting for no reward when the bad-news stimulus is presented. From Equation 1, R_{Info} becomes

$$R_{Info} = \frac{p_{Info}}{s + p_{Info} \times (t+h) + \cancel{(1 - p_{Info}) \times t}} = \frac{1}{\frac{s}{p_{Info}} + t + h} \tag{2}$$

where p_{Info} refers to the probability of reward in the Informative Option; because this option only provides reinforcement when the good-news stimulus is presented, p_{Info} is the same as the probability of presenting the good-news stimulus and 1- p_{Info} is the same as the probability of presenting the bad-news stimulus. Because the probability of bad-news (1-p_{Info}) is absent in Equation 2, it has no effect on the perceived rate intake in the Informative Option, so it should not affect preference.

On the other hand, because in the Non-informative Option the animal never knows whether food will come or not, it actively waits. Thus, the rate of intake in the Non-informative Option, $R_{Non\text{-}info}$ remains

$$R_{Non-info} = \frac{p_{Non-info}}{s + p_{Non-info} \times (t+h) + (1 - p_{Non-info}) \times t} = \frac{1}{\dfrac{s+t}{p_{Non-info}} + h} \tag{3}$$

Figure 2b shows the same information as Figure 2a, using the new rate functions (Equations 2 and 3). Note that in Figures 2a and 2b, the rate function for the Non-informative Option is the same, because in both cases all durations are taken into account; the difference in the rate functions for the Informative Option is that in Figure 2a the bad-news durations are included whereas in Figure 2b they are not. Notably, the new functions predict again that animals should prefer the Non-informative Option for the probabilities used in Vasconcelos et al.'s (2015) study.

The derivation of Equations 2 and 3 from optimal foraging theory include the ITI (or the time spent searching in the natural environment, s) in the rate of each option. However, there is an important difference between the ITI in the laboratory and the searching time in the natural environment: While in the natural environment the searching time occurs *after* the animal decides to start searching, and thus is included in the rate of the chosen option, in the laboratory the ITI occurs *before* the animal decides to start the trial (and after the ultimate event – the reward). In such cases, it has been shown that the ITI may not be attributed to a specific option and is thus unlikely to influence preference (Bateson & Kacelnik, 1996; Logue, Smith, & Rachlin, 1985; Mazur, 1989; Spetch, Belke, Barnet, Dunn, & Pierce, 1990; Spetch et al., 1994; Stephens & Anderson, 2001). Figure 2c shows that, if the ITI is dropped from the rate functions ($s = 0$ in Equations 2 and 3), the model predicts a preference for the Informative Option, except when $p_{\text{Info}} = 0$ or $p_{\text{Non-info}} = 1$, which is consistent with previous studies (e.g., Mazur, 1996; Stagner et al., 2012; Vasconcelos et al., 2015; Zentall et al., 2015). In other words, with a constant probability of reinforcement in the Non-informative Option, manipulating the probability of the bad and good-news stimuli in the Informative Option should not affect preference for the Informative Option, unless the bad-news stimulus is always presented.

Besides predicting that the probability of the bad-news stimulus should not affect preference for the Informative Option (except when this stimulus is always presented), another counterintuitive prediction derived from Equation 2 is that the time spent in the presence of the bad-news stimulus should also be irrelevant: By eliminating the bad-new trials from Equation 2, the time t spent in the presence of the bad-news stimulus is eliminated.

The aim of this paper was to test the two aforementioned predictions, namely that (a) a preference for the Informative Option should always be observed independently of the values of p_{Info} except when $p_{\text{Info}} = 0$; we tested this predictions by manipulating the probability of the bad-news stimulus (Experiment 1), and; (b) increasing the duration of the bad-news stimulus should not affect preference because this duration is not included in the rate function (Experiment 2).

2. Experiment 1 – Probability of “bad news”

In this experiment, we increased the probability of bad news in the Informative Option from .80 to 1.00 (and therefore decreased the probability of good news from .20 to 0.00). The probability of food following each nondiscriminative stimulus in the Non-informative Option was held constant at .5. Our goal was twofold: (1) Confirm Vasconcelos et al.’s (2015) findings that the probability of bad news bear little to no weight on choice, but using a counterbalanced design. In the original findings, the probability of bad news was manipulated in an ascending order which does not exclude the hypothesis that resistance to change may have played an important role; (2) track the change in preference for the Informative Option with a more fine-grained distribution of parameters. In Vasconcelos et al.’s (2015) study, preference decreased abruptly from indifference to virtually zero when the probability of bad news increased from .95 to 1.00. We aimed at sampling the underlying preference function with more parameter values.

2.1. Method

2.1.1. Subjects

Six pigeons (*Columba livia)*, maintained at about 80% of their free-feeding weights participated in this experiment. The birds had experience with a variety of procedures (temporal and numerosity discriminations), but none had experience with this task. In the colony room, all birds were housed individually and had water and grit continuously available. The colony room had controlled temperature (about 21ºC) and a 13h:11h light/dark cycle with lights on at 8 AM.

2.1.2. Apparatus

Three operant boxes for pigeons from Med Associates were used. The boxes were 28.5-cm high, 24-cm long and 30-cm wide. Each box was enclosed in a sound-attenuating chamber, equipped with a ventilation fan that also masked outside noises. The response panel had three centrally aligned keys, 6 cm apart, center-to-center. The keys were circular with 2.5 cm in diameter, and the lowest edge was 21 cm above the floor grid. Below the response keys, centrally aligned and 4 cm above the floor grid there was a grain hopper opening (6-cm wide x 5-cm high). When activated, the food hopper was illuminated with a 1.1-W light. In the wall opposite to the response panel, centrally located and 23 cm above the floor there was a houselight (2.8 W) that illuminated the whole box. A personal computer with a custom Visual Basic 2010 program controlled the events and recorded data via Whisker software (Cardinal & Aitken, 2010).

2.1.3. Procedure

Pretraining. All birds were trained to peck each of the stimuli used in this task: red, green, yellow, blue and white on the left and right keys, and white on the center key. Each session consisted of 44 trials, 4 trials with each stimulus. On each trial, one randomly selected stimulus turned on, and after one peck food was available for 3 s. A 10-s ITI with the houselight on separated the trials. In the second pretraining session, the peck requirement was increased to 10 pecks.

Training. Each session comprised 120 trials, 80 forced-choice trials (40 with each option) and 40 free-choice trials. The free- and forced-choice trials were randomly interspersed, with the constraint that in each set of 30 trials, 20 were forced-choice (10 with each option) and 10 were free-choice trials.

At the beginning of a choice trial, the white center key started flashing (250 ms on, 250 ms off). One peck at the center key turned it off and illuminated the side keys (both white hues, see Figure 1), starting the initial links. For each bird the Informative and the Non-informative options were always presented on the same side, but side allocation was counterbalanced across pigeons.

If the pigeon pecked the Informative Option key, the Non-informative Option key turned off, and with probability *p* the bad-news stimulus (e.g., green) was turned on. After 10 s the terminal-link stimulus was turned off and a 10-s ITI with only the houselight on followed. Pecks during the terminal links were recorded but had no programmed consequence. On the remaining 1 - *p* proportion of trials, the good-news stimulus (e.g., red) was turned on for 10 s, food was delivered immediately after, and the illuminated ITI followed.

If the pigeon instead pecked the Non-informative Option key, the Informative Option key was turned off, and on 20% of the trials one terminal stimulus (e.g., yellow) was turned on; on the remaining 80% of the trials the other terminal stimulus (e.g., blue) was turned on. In both non-informative terminal links, the key remained lit for 10 s, food was delivered on a randomly selected half on the trials, and the 10-s ITI followed. The terminal-link hues were counterbalanced across pigeons, with the restriction that red and green were always associated with one option and yellow and blue were always associated with the other option. The reinforcement duration varied from bird to bird (between 2.5 and 6 s) in order to avoid feeding outside the experimental chamber.

The forced-choice trials were structurally similar to free-choice trials, except that after a peck to the flashing center key only one of the side keys, the Informative or the Non-informative Option, was turned on.

If a peck to the center key (to start the trial) or to the side keys (initial-link phase) did not occur within 15 s since the key illumination, the trial ended, a 2-minute timeout followed with all lights off, and the trial was repeated. After about 10 sessions, the timeouts occurred rarely (on less than 1% of the trials).

The probability of bad news, *p*, varied from 0.80 to 1.00 according to Table 1. The first three conditions experienced by each bird lasted for 20 sessions, and the last two conditions lasted for 30 sessions. In each condition, the mean proportion of choices during the last three sessions was used as the measure of preference. A significance level of 0.05 was used in all analyses.

Table 1. Individual preferences for the Informative Option, for each probability condition. Numbers in parenthesis show the order of conditions for each bird.

Condition	Pigeon						Mean
	P236	P604	P229	P389	P230	P751	
p = .80	0.97 (1)	1.00 (3)	1.00 (4)	1.00 (2)	1.00 (2)	1.00 (1)	0.99
p = .90	0.68 (3)	1.00 (1)	1.00 (2)	1.00 (4)	0.99 (4)	0.99 (3)	0.94
p = .95	0.23 (4)	1.00 (2)	0.99 (1)	0.99 (3)	0.96 (1)	0.86 (2)	0.84
p = .975	0.11 (2)	1.00 (4)	0.69 (3)	0.06 (1)	0.50 (3)	0.89 (4)	0.54
p = 1.00	0.00 (5)	0.00 (5)	0.01 (5)	0.01 (5)	0.05 (5)	0.01 (5)	0.02

2.2. Results and Discussion

Figure 3 shows the average preference for the Informative Option (± 1 SEM) as a function of the probability of the bad-news stimulus. When the probability of bad news was 0.80, we replicated previous results, with pigeons strongly preferring the suboptimal, Informative Option (M = .99, SEM = 0.01). As the probability of bad news increased, preference for the Informative Option progressively decreased. This decrease was confirmed by a repeated measures analysis-of-variance (ANOVA), revealed by a significant effect of probability, $F(4, 20) = 24.76$, $p < .001$, $\eta_G^2 = 0.75$. Overall, preference was significantly above chance when p = .80, p = .90 and p = .95 (one-sample t-tests: $t(5) = 89.00$, $p < .001$, d = 36.33, 95% CI for d [18.85, 53.82]; t (5) = 8.26, p <. 001, d = 3.37, 95% CI [1.33, 5.41]; $t(5) = 2.71$, $p = .04$, d = 1.11, 95% CI [-0.24, 2.45], respectively), and was significantly below chance when $p = 1.00$ ($t(5) = -58.00$, $p < .001$, d = 23.68, 95% CI [12.24, 35.11]). When p = .975, animals were, on average, indifferent between the options ($t(5) = 0.26$, $p = .81$, d = 0.11, 95% CI [-1.13, 1.35]).

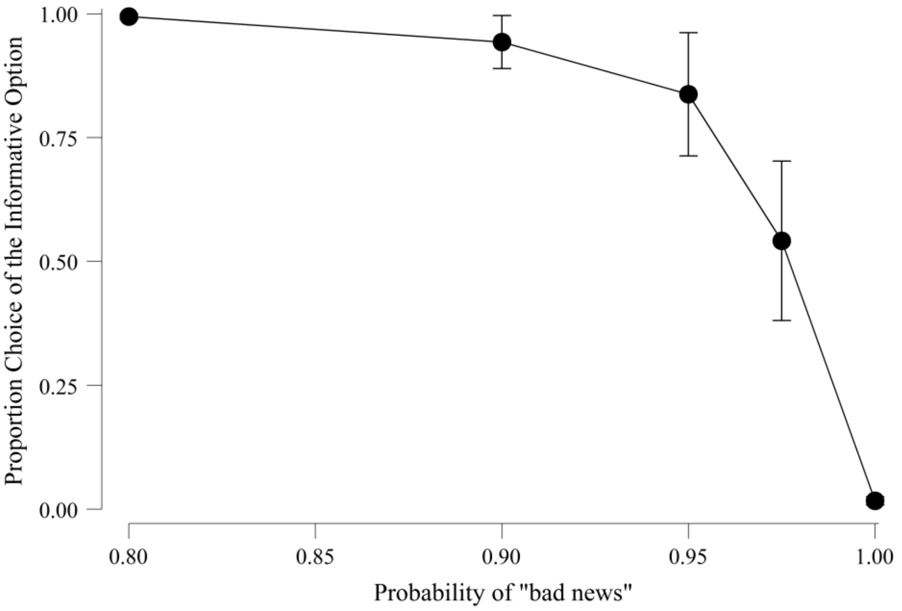


Figure 3. Mean proportion of choices of the Informative Option in the last three sessions of each condition in Experiment 1. The error bars are the standard error of the mean.

Although Figure 3 shows that the higher the probability of bad news, the less animals preferred the Informative Option, this continuously decreasing function is highly influenced by the results of one animal, P236 (first data column in Table 1). When the probability of bad news increased from 0.80 to 0.95, the decrease in preference was mostly due to this bird. More specifically, when p = .90, five birds chose suboptimally at least 99% of the time, whereas for P236 preference was 67.5%. Similarly, when p = .95, the same five birds chose suboptimally at least 86% of the time, but preference for P236 was already at 22.5%. Thus, increasing the probability of bad news from 0.80 to 0.95 had only a small effect on preference, except for P236. It is worth noting that pigeon P236 was the bird that had the greatest increase in the probability of bad news: from 0.80 in the first condition to 0.975 in the second condition. Possibly, after experiencing this hefty change, this bird would have needed more sessions than the others to recover preference for the Informative Option in subsequent conditions.

When the probability of bad news was 0.975, a maximum of two rewards per session were delivered in the Informative Option: one on forced-choice trials and possibly one more on a free-choice trial (the obtained percent of reinforcement on the Informative Option varied between 2.2% and 2.8% across pigeons, M = 2.5%). Despite the fact that the Informative Option was rarely reinforced in this condition, there was great variability between subjects: three preferred the Informative Option significantly above chance (P604, P229 and P751;

binomial test *p* < .001), two preferred the Non-informative Option (P236 and P389; binomial *p* < .001) and P230 was indifferent between the two options (binomial *p* = .54). A closer look at the data revealed important order effects (see Figure 4 and Table 1 for individual data). When the condition with p = .975 was either the first or the second condition (as for P389 and P236), preference for the Informative Option was clearly below chance. Moreover, the later in training p = .975 was in effect, the higher was the preference for the Informative Option.

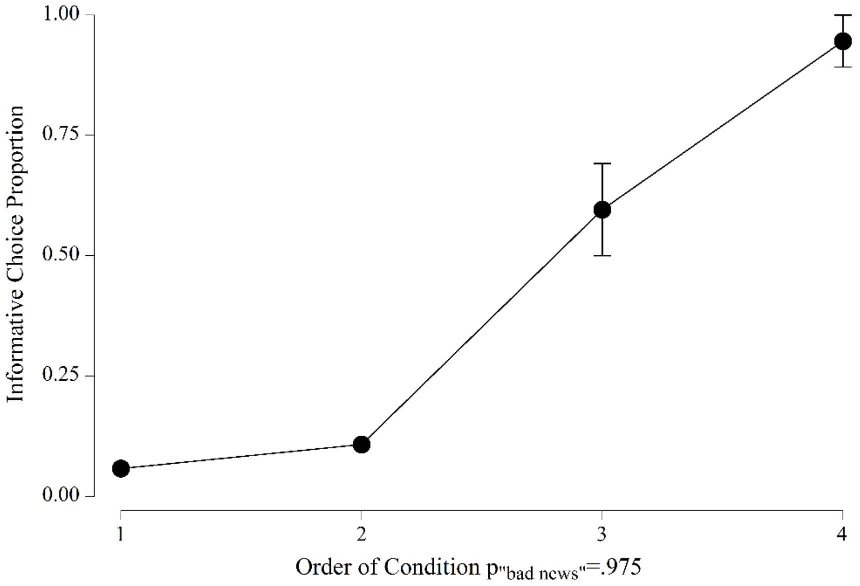


Figure 4. Mean proportion of choices of the Informative Option in the last three sessions of Condition p = .975, as a function of order of presentation of that condition (cf. Table 1). For orders 1 and 2 data refers to only one bird, and for orders 3 and 4 data refers to two birds. The error bars are the standard error of the mean.

Because the first condition ended after 20 sessions, regardless of the probability in effect, the pigeon that started the experiment with p = .975, P389, had less contact with the good-news stimulus in this condition compared to the other pigeons that started with higher probabilities of good news. Thus, one could argue that the discrimination between the good and bad-news stimuli was not well established for P389, and that is why preference for the Informative Option did not develop. One way to measure discrimination between two stimuli is by calculating the discrimination ratio, that is, the responses made in the presence of the positive stimulus (i.e., the good news), divided by the sum of responses to the positive and negative stimuli combined (i.e., the good and bad-news stimuli). A perfect discrimination would yield a ratio of 1, meaning that the bird only pecked the positive stimulus. The average discrimination ratio during the last three sessions of Condition p = .975, from order one to four was 0.81, 1.00, 0.91 and 0.88, respectively. Therefore, all pigeons were discriminating the good and bad news stimuli when p = .975, regardless of whether it was the first condition or not. The increase in preference for the suboptimal alternative as the Condition p = .975 occurred later in training may indicate a resistance to change due to the experimental history.

In summary, the results show that animals preferred the Informative Option even when the Non-informative Option provided 10 times more food (50% vs. 5% of reinforced trials). These results are even stronger than those reported by Vasconcelos et al. (2015), as their starlings were already indifferent between options when p = .95. Currently, it is unclear whether this difference is reliable and, in the affirmative, whether it is due to different experimental histories, procedural details (e.g., ascending vs. counterbalanced p) or even to species differences. Moreover, although the average data shows a modulation of the preference for the Informative Option as a function of the probability of bad news (Figure 3), this result was mostly influenced by the results of one animal (P236), and by the order effects shown in condition p = .975. If we excluded data of P236, preference for the Informative Option in conditions p = .80, p = .90 and p = .95 would be 100%, 100%, 96%, respectively, so one would conclude that there was little or no effect of these conditions. Then, when p = .975 preference would be 63%, but the order effects obtained in this condition make this preference difficult to interpret. Finally, when p = 1.0, preference decreased to 2%. Altogether, these results are consistent with the predictions of Vasconcelos et al.'s (2015) model. Our results also show that the reinforcement history and the transition between probabilities may be important factors to take into account in these experiments. To control for order effects, a between-subjects design would be appropriate.

3. Experiment 2 – Duration of "bad news"

In this experiment, we manipulated the duration of the bad-news stimulus (hereafter the bad-news delay). Recall that one of the predictions of Vasconcelos et al.'s (2015) model is that the duration of this stimulus ought to be irrelevant vis-à-vis preference. This is a highly counterintuitive hypothesis and a strong test of the model. To test whether the prediction holds and study its boundary conditions, we used a titration procedure: When pigeons preferred the Informative Option, we increased the duration of the bad-news stimulus; when they preferred the Non-informative Option we decreased the duration of the bad-news stimulus.

3.1. Method

3.1.1. Subjects

Eight pigeons (*Columba livia)* participated in this experiment. All birds were maintained as described in Experiment 1. Previous to this experiment, all birds participated in a choice experiment in which they acquired an exclusive preference for one of the side keys. To make sure that pigeons would not show, in the present experiment, a preference for the Informative Option due to a preexisting side bias, for each bird, the Informative Option was assigned to the non-preferred side in the previous experiment. If anything, this assignment would produce a bias *against* the Informative Option.

3.1.2. Apparatus

Four operant boxes for pigeons were used: the three from Experiment 1 and a similar, custom-built box. This box was 31-cm high, 33-cm long and 33-cm wide. The box was enclosed in a sound-attenuating chamber, equipped with a ventilation fan that also masked outside noises. The response panel had three centrally aligned keys, 9 cm apart, center-to-center. The keys were circular with 2.5 cm in diameter, and the lowest edge was 21 cm above the floor grid. Below the response keys, centrally aligned and 6.5 cm above the floor grid there was a grain hopper opening (6-cm long x 4.5-cm high). When activated, the food hopper was illuminated with a 1.1-W light. In the wall opposite to the response panel, centrally located and 27.5 cm above the floor, a houselight (2.8 W) provided general illumination.

3.1.3. Procedure

The general procedure was the same as in Experiment 1, with the following exceptions. First, the probability of the bad-news stimulus was kept constant at 80%. Second, there was no limit on the time to peck the center and side keys, so there were no timeouts. Finally, the experiment had two conditions: firstly an adjusting-delay procedure was implemented for the bad-news delay, and then the bad-news delay was fixed, as described below.

Pretraining. The pretraining was the same as in Experiment 1.

Adjusting-delay Condition. Each session was divided into blocks of 30 trials. Within each block, the first 20 trials were forced-choice trials (10 with the Non-informative Option and 10 with the Informative Option, interspersed), and the last 10 trials were free-choice trials. After each 30-trial block, the bad-news delay could increase, decrease, or stay the same depending on the pigeon's choices in the previous block. If the pigeon chose the Informative Option on two trials or less, the delay decreased by 5 s; if it chose the Informative Option on eight trials or more, the delay increased by 5 s; otherwise, the delay remained the same as in the previous block.

In the first block of the first session the bad-news delay was set at 10 s. In the following sessions, the bad-news delay on the first block of trials was computed for each bird based on its choices in the last block of the previous session. The lower limit for the bad-news duration was 0 s. In that case, if the pigeon chose the Informative Option, the good-news stimulus appeared for 10 s on 20% of the trials and on the remaining 80% of the trials the choice peck was immediately followed by the ITI. Although we did not plan to include an upper limit for the bad-news delay, as the experiment proceeded it became necessary to do so. Therefore, the maximum delay that could be set at the beginning of a session was 200 s even though it could surpass 200 s within a session. For example, if, in the last block of a session, the delay was 195 s and the bird had chosen the Informative Option eight or more times, the delay at the beginning of the last session was set to 200 s but it could increase to 205 s and 210 s if the bird continued to show a preference for the Informative Option. This condition was run until the bad-news delay was considered stable or reached 200 s, whichever came first.

Because increasing the bad-news delay made the sessions longer, we decreased the number of 30-trial blocks per session as the delay increased. If the delay at the beginning of a session was equal to or shorter than 120 s, the session comprised five blocks (150 trials); if the delay was between 120 s and 180 s inclusive, the session comprised four blocks (120 trials); and if the delay was longer than 180 s, the session comprised three blocks (90 trials).

Fixed-delay Condition. To evaluate whether preference for the Informative Option could be caused by the adjustment of the bad-news delay, we ended the experiment with three sessions in which the bad-news delay was kept constant at the value reached in the Adjusting-delay Condition or at 200 s in case it reached 200 s or more in the previous session. Each session comprised three 30-trial blocks (90 trials).

3.2. Results and Discussion

Adjusting-delay Condition. All birds started the experiment by preferring the optimal, Non-informative Option, reaching exclusive preference in the first session. This preference can be explained by the fact that the Non-informative Option was assigned to the preferred key in their previous experiment. As a consequence of choosing the Non-informative Option, during the first session the bad-news delay decreased to 0 seconds. Of interest was whether the bias for the Non-informative Option would attenuate and pigeons would start choosing the Informative Option, and continue choosing it when the bad-news delay started to increase. After several sessions, pigeons started to choose the Informative Option eight or more times in each block (of 10 free-choice trials), so the bad-news delay started to increase (see Figure 5).

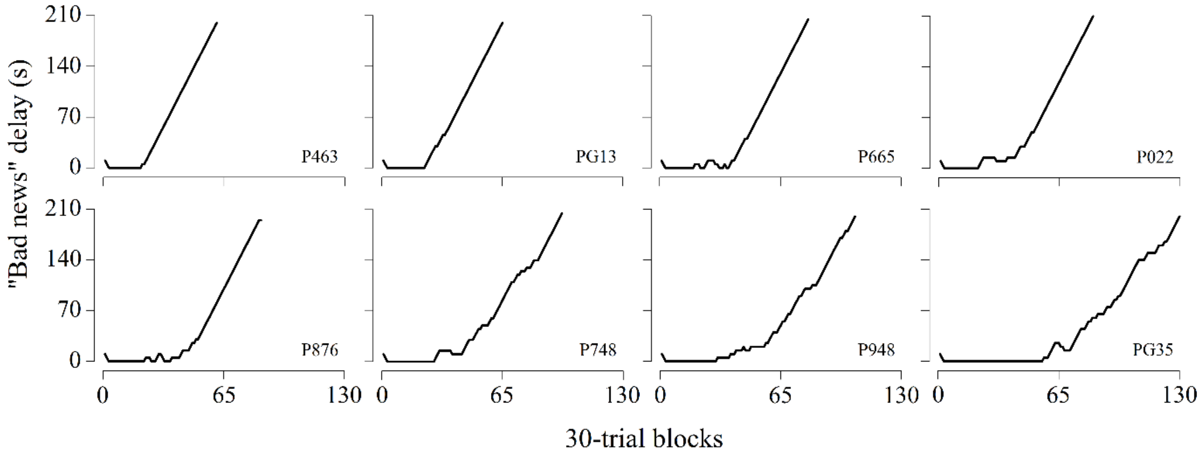


Figure 5. Bad-news duration (in seconds) in each 30-trial block, during the Adjusting-delay Condition.

All birds continued to choose the Informative Option on more than 80% of the choice trials, and the bad-news delay continued to increase, reaching at least 200 s in one session. Due to an experimenter error, pigeon P876 ended the Adjusting-delay Condition when it reached 195 s. For the other birds, the maximum delay during the Adjusting-delay Condition was 200 s for birds P463, PG13, P948 and PG35, 205 s for P665 and P748 and 210 s for P022. The fact that the birds reached the maximum delay suggests that the duration of the bad-news stimulus has no effect on the value of the Informative Option (the left panel of Figure 6 shows the average preference for the Informative Option, ± 1 SEM, during the last three blocks of this condition).

During the first session, the bad-news duration had decreased to 0 s. Therefore, when the Informative Option was chosen, the pigeons experienced the good-news stimulus on 20% of the trials, and the immediate onset of the ITI on the remaining 80% of the trials. At this point, animals started preferring the Informative Option. There are possibly two reasons for animals to start choosing more the Informative Option. First, animals may have stopped paying attention to the ITI because, as the bad-news stimulus, it was not associated with food. As a consequence, the Informative Option was perceived as an option that always delivered food. Second, animals may have continued to pay attention and actively wait in the presence of the Non-informative stimuli, paying the cost of not receiving food on half of the trials. As a result, animals started preferring the Informative Option, thus increasing the duration of the bad-news stimulus. Interestingly, on those first occasions with the bad-news stimulus present, the pigeons pecked at it as much as they pecked at the good-news stimulus (discrimination ratio close to 0.5). With additional exposure to the bad-news stimulus, the pigeons ceased pecking at it and the discrimination ratio increased to 1.0. This means that, for a suboptimal preference to develop, it is not necessary to have a specific stimulus associated with the absence of food in the Informative Option, because preference can develop when the bad news is the ITI, which occurs at the end of all trials.

Fixed-delay Condition. The right panel of Figure 6 shows preference for the Informative Option, ± 1 SEM, during the Fixed-delay Condition.

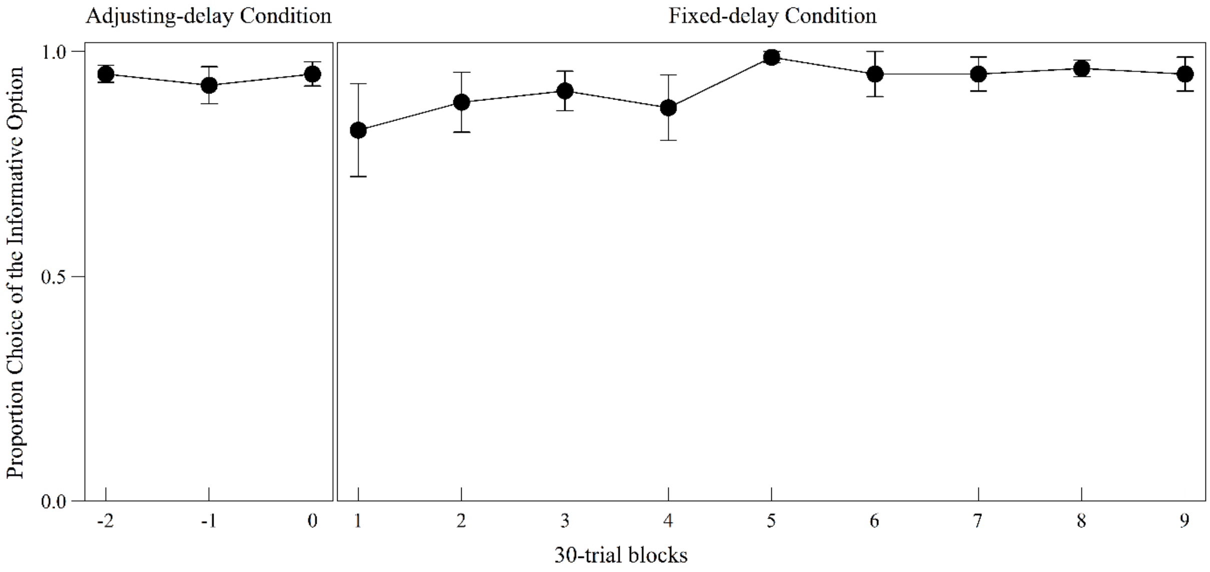


Figure 6. Mean proportion of choices of the Informative Option in the last three blocks of the Adjusting-delay Condition and during the nine blocks of the Fixed-delay Condition, in which the bad-news delay was fixed at 200 s for all birds.

In the last block of the Adjusting-delay Condition all birds were choosing the Informative Option in at least 80% of the trials (left panel of Figure 6). When the bad-news delay was fixed at 200 s, one bird showed a substantial decrease in preference (PG35; see individual data in the Appendix A). The reasons for the decrease are unclear, given that the delay (and all the other parameters) remained the same as in the last block of the previous session. Note, however, that this bird had the most variability in performance during both conditions, so the decrease in preference between conditions should be interpreted with caution.

Even though some birds decreased preference for the Informative Option during the first block, by the last block (last data point in Figure 6) all of them were choosing the Informative Option reliably above chance (M = .95, SEM = 0.04, $t(7) = 11.91$, $p < 0.01$, d = 4.21, 95% CI [2.24, 6.18]). In fact, preference during the last three sessions of the Fixed-delay Condition and the last three sessions of the Adjusting-delay Condition did not differ significantly: A repeated-measures ANOVA with condition (two levels) and session (three levels) as within-subjects factors revealed no significant effect of condition, $F(1, 14) = .264$, $p = .623$, $\eta^2_G.\ = 0.006$.

In summary, this experiment showed that pigeons strongly prefer an option that provides information about the delivery of food, even when 80% of the trials with such option are spent in the presence of a signal for the absence of food for more than 3 minutes.

This fact indicates that the bad-news stimulus does not function as a punisher or as a conditioned inhibitor, in the sense that it does not seem to affect preference for the Informative Option. If we estimate that the peck to start the trial took about 2 s and the peck on the initial link took 2 s as well, and include the 10-s ITI and the terminal-link durations on each key, then the preferred, Informative Option yielded about 0.07 reinforcements per minute, whereas the Non-informative Option yielded 1.25, over 18 times more.

4. General Discussion

Several studies have shown that animals prefer signaled to unsignaled rewards (Bower et al., 1966; Prokasy, 1956; Roper & Zentall, 1999). Moreover, this preference even holds when the signaled option is less profitable than the unsignaled one (for a review, see Zentall, 2014, 2016). In particular, this preference for the suboptimal option occurs when (a) the negative stimulus (i.e., the bad-news stimulus in the Informative Option) signals absence of food, and (b) the good-news stimulus in the Informative Option is a better predictor of food than the stimuli in the Non-informative Option. To account for this ostensively paradoxical preference for an option that, although informative, yields a lower rate of reinforcement, researchers have theorized that animals do not take into account the negative stimulus, and behave as if they were in a situation without bad news. The model developed by Vasconcelos and colleagues (2015) gives center stage to this assumption. Based on a functional analysis, they considered that forcing an animal to wait in the presence of a signal for no food (as the bad-news stimulus) may not be common in natural settings, so it is likely that in such situations animals will use a strategy that is adaptive in the natural environment - do not take into account the bad-news stimulus. Assuming that animals always use the strategy of ignoring the bad-news stimulus, they advanced two counterintuitive predictions: (1) Preference for the Informative Option should always be observed regardless of the values of p_{Info}, except when $p_{\text{Info}}= 0$, and; (2) the duration of the bad-news stimulus should not affect preference because it is not included in the rate functions. We tested these two predictions in Experiments 1 and 2, respectively.

In Experiment 1, we manipulated the probability of bad news, increasing it from 0.80 to 1.00. By doing so, we decreased the overall percentage of reinforced trials in the Informative Option from 20% to 0%, while keeping it at 50% in the Non-informative Option. We found that animals preferred the Informative Option even when the probability of reinforcement in the other alternative was 10 times higher (5% *vs.* 50%). These results are similar to, but more extreme than those reported by Vasconcelos et al. (2015): Both starlings and pigeons showed preferences of about 100% for probabilities of bad news equal and below 0.90, but whereas starlings' preference decreased to chance levels when the probability of bad news was set at 0.95, our pigeons' preference decreased to only 85%, remaining significantly above chance.

These results are also consistent with Mazur's findings using an adjusting-delay procedure. In one condition, Mazur (1996) gave pigeons a choice between two delayed rewards: one certain and the other probabilistic. In the probabilistic option, with probability p the key was immediately turned off and an ITI followed (which can be seen as a signal for no food), and with probability 1-p the key stayed on for 20 s and was always followed by food. In the certain option, the delay was increased and decreased until the birds were indifferent between the two alternatives. Mazur found that, at indifference, the delay in the certain option was about 20 s, regardless of the probability of the key turning off – and not giving food – from 20% to 80%. That is, it seems that the animals were ignoring the no-food trials, thus making the probabilistic option functionally similar to the certain option, which resulted in the indifference between options revolving around the same delay value: 20 s.

Mazur (1996) did not test a probability of bad news higher than 0.80, and Vasconcelos et al. (2015) tested with 0.95 and 1.00. We tested an intermediate value, 0.975, and obtained mixed results. Overall, the longer the history of choosing the Informative Option, the more likely it was that the birds continued to choose this option even with such a low probability of reinforcement. It remains possible that with more sessions of training the preference for the Informative Option would be stronger, even with a very low probability of reinforcement.

The fact that pigeons seem to reduce their preference for the Informative Option even before the probability of bad news is set to zero (see the concave function in Figure 3) may mean that: (a) instead of choosing the option yielding the higher perceived rate of reinforcement, the choice algorithm may be more complex than previously assumed, or (b) with such high probabilities of bad news the rewarded trials are extremely rare and may therefore bear little weight if animals used, for instance, a moving window to compute rates. Note, however, that if we exclude the results of one animal that was clearly more affected by the probability of bad news than the other birds (P236), and interpreting with caution the results when p = .975, preference for the Informative Option remained high and only reversed when p = 1, as predicted by Vasconcelos et al.'s (2015) model.

In Experiment 2, we used a titration procedure to manipulate the duration of the bad-news stimulus: When animals preferred the Non-informative Option the bad-news duration decreased; when they preferred the Informative Option that duration increased. According to Vasconcelos et al. (2015), animals should keep preferring the Informative option regardless of the bad-news delay. We found that all eight birds continued to choose the Informative Option (95% of the time, on average) when the bad-news duration was 200 s, that is, when the reinforcement rate was roughly 18 times lower than on the Non-informative Option. This result strongly supports the hypothesis that animals ignore the time they spend in the presence of a stimulus that signals absence of food.

Previous studies (Mazur, 1995; Spetch et al., 1994; Pisklak, McDevitt, Dunn, & Spetch, 2015) have also found results pointing in this direction, but they tested much shorter durations than we did. For instance, Spetch et al. (1994) increased the duration of the bad-news stimulus from 5 s to 55 s. Although the average preference for the Informative Option decreased when the bad-news stimulus duration increased, this decrease was not statistically significant.

One could argue that the bad-news stimulus did not affect choice in Experiment 2 because pigeons avoided being in its presence (turning around, for example). In fact, in the present studies we observed that the birds usually turned their back to the keys when the bad-news stimulus appeared. To test this avoidance hypothesis, Stagner, Laude and Zentall (2011) replaced the bad-news stimulus by a houselight that would be difficult to avoid. The preference for the Informative Option was the same as when the bad news was a keylight (as in the present experiments). Thus, even when the bad-news stimulus is difficult to avoid, pigeons still choose suboptimally. It is thus unlikely that avoidance of the negative discriminative stimulus *per se* can account for the suboptimal preference we observed.

Interestingly, the model developed by Vasconcelos and colleagues (2015) based on an optimality analysis has noticeable affinities with other models emphasizing proximate causes. Mazur (1995, 1996) and Dunn and Spetch (1990), for instance, developed different accounts emphasizing the relevance of conditioned reinforcers. Even though these latter accounts do not always make the same predictions, both assume that a stimulus that is never followed by food is ignored. Mazur's hyperbolic discounting model states that the present value of a delayed reward decreases hyperbolically as the delay to reward increases. However, only delays in the presence of conditioned reinforcers (i.e., stimuli associated with food) are taken into account. This means that when a stimulus is never followed by food it should not be considered and it should not affect the option's value. Thus, Mazur's (1995, 1996) model predicts that the value of the Informative Option in our study has the same value as an option that always delivers food after 10 s. In a series of experiments, Mazur's predictions were usually confirmed (Mazur, 1989, 1991, 1995, 1996).

A different account for preference for the Informative Option (Dunn & Spetch, 1990) comes from a modified version of the Delay-reduction theory (Fantino, 1969; see also Ward, Gallistel, & Balsam, 2013 for a similar idea). The core of Dunn and Spetch's (1990) account lies on the idea that a signal has greater conditioned-reinforcement value when it signals a greater reduction in the delay to reinforcement. Specifically, using the task depicted in Figure 1, when the animal chooses the Informative Option, the expected delay to food is 50 s. When it sees the good-news stimulus, that cue signals a large reduction in the delay to food (from 50 s to 10 s) so the good-news stimulus becomes a strong conditioned reinforcer; if, on the other hand, it sees bad-news stimulus, no food is given and therefore this stimulus does not become a conditioned reinforcer. When the Non-informative Option is chosen the expected delay to food is 20 s, and when the terminal-link stimuli appear, they do not signal a reduction in delay to food (the expected delay remains 20 s) and because of that they are not considered strong conditioned reinforcers. In other words, the Non-informative Option provides delayed primary reinforcement, whereas the Informative Option provides delayed primary reinforcement *and* an immediate and strong conditioned reinforcement. As in Mazur's (1995, 1996) account, Dunn and Spetch (1990) framework predicts no effect for manipulations on the bad-news stimulus. Congruent with both accounts, our results showed that manipulations of duration of bad news had nearly no effect on choice, and manipulations of probability of bad news had only a significant effect when the probability of reinforcement in the Informative Option was very low.

Even though both Mazur's (1995, 1996) and Dunn and Spetch's (1990) accounts correctly predict that the bad news should have no effect on choice, they did not advance an explanation for *why* animals ignore the bad news. Vasconcelos et al. (2015) went that step forward. The model is based on the animals' ecology and evolution (Stephens, Brown, & Ydenberg, 2007), and has three primary premises. First, animals maximize their (perceived) rate of food intake. Second, preference depends on the ratio of reinforcement rate of one option to the reinforcement rate of the other available option. Third, not all durations are considered in the computation of the reinforcement rate: The time spent in the presence of stimuli that are never followed by food (during the bad news and the ITI) do not enter into the rate calculations. The main reason for not taking these durations into account is that, outside the laboratory, animals would not pay the cost of waiting in the presence of those stimuli, but would rather leave to search for a patch with a higher probability of food. In terms of the psychological mechanism, one could say that, in the presence of a stimulus associated with the absence of food, there is no conditioning with that stimulus. Without conditioning the animal does not pay attention to the stimulus, that is, does not associate the costs of waiting in the presence of this stimulus with the choice previously made (in the Informative Option). Therefore, the animal does not devalue the Informative Option, even though the bad-news stimulus is presented frequently. In the Non-informative Option, because the signals are associated with food (as the good-news stimulus), there is conditioning and the animal pays the cost of half of the times waiting and not getting food. As a consequence, the Informative Option is perceived as a more profitable option, and is preferred.

This is another example of an irrational behavior that arises from the mismatch between the experimental environment and the environment in which the behavioral mechanism evolved (the problem of the domain of testing vs. the domain of selection; see, for example, Stevens & Stephens, 2010). In other words, the so-called “suboptimal” choice is basically a consequence of using a mechanism that is advantageous in the wild but is suboptimal in a controlled, artificial environment like the operant box where the experiments are run. It is assumed nowadays that, as animals’ bodies evolved through natural selection, the same happened with their behavior and cognition. We can thus try to understand behavior in light of the selective pressures and the evolutionary environment in which it supposedly evolved. Such an analysis frequently reveals that some apparently irrational behavior is in fact adaptive under certain circumstances.

CHAPTER IV

STUDY 2:

DO PIGEONS ESCAPE FROM "BAD NEWS"?

Abstract

Animals evolved to use the information available in the environment. When, in the natural environment, an animal encounters a stimulus that signals the absence of food – a ‘bad-news’ stimulus – it will most likely use that information and spatially redirect its search for another prey or patch. Because the animal never pays the cost of waiting in the presence of a bad-news stimulus, the duration of such stimulus played no role in the evolution of the decision processes deployed in these circumstances. When, in the laboratory, animals are forced to experience the duration of the bad-news stimulus, they seem to *ignore* the imposed bad-news delay as they would do in the natural environment, even though they cannot redirect search as they would do in the wild. Under certain circumstances, this behavior can lead to suboptimal preferences, such as a preference for an option yielding a low rather than a high rate of reinforcement. In the present experiments, we tested in pigeons the Vasconcelos, Monteiro and Kacelnik’s (2015) prediction that, if animals are given the opportunity to escape the bad-news stimulus, they will do so. We adapted the prey choice model to predict when the escape response should occur. Consistent with our predictions, we found that the shorter the intertrial interval and the longer the duration of the signals predicting food (or no food), the higher the proportion of escape responses. Based on these results, we discuss the meaning of *ignoring* bad news.

Keywords: suboptimal choice, bad news, escape response, prey choice model, pigeons

1. Introduction

Only a few things are certain in life. In an uncertain environment, where a particular response does not always produce the same outcome or a stimulus does not always predict the same result, reliable information about important events is highly valuable. Information that reduces uncertainty about food delivery is undoubtedly important for animals to the point that, under certain circumstances, animals will forego food in order to secure information.

For example, Gipson, Alessandri, Miller and Zentall (2009) showed that pigeons prefer an Informative Option that signals immediately whether food will occur or not, but yields less food on average, over a Non-informative Option that does not signal whether food will follow, but provides more food on average. In their experiment, after pigeons chose one of two possible options, a stimulus appeared for 30 s. When they chose the Informative Option, on half of the trials, a “good-news” stimulus (i.e., a positive discriminative stimulus), which was always followed by food, was presented; on the remaining half of these trials, they saw a “bad-news” stimulus (i.e., a negative discriminative stimulus) that was never followed by food. If, instead, they chose the Non-informative Option, one of two equiprobable stimulus was presented, both ending with food 75% of the time. Even though the overall probability of food was higher in the Non-informative Option than in the Informative Option (.75 vs. .50), pigeons reliably chose the latter. In this task pigeons behaved suboptimally in the sense that they preferred the learner option. An even stronger suboptimal preference was found by Stagner and Zentall (2010) when they decreased the overall probability of reinforcement in the Informative and Non-informative options to 20% and 50%, respectively (see also Stagner, Laude, & Zentall, 2011; Vasconcelos, Monteiro, & Kacelnik, 2015; Zentall & Stagner, 2011) Figure illustrates the general procedure.

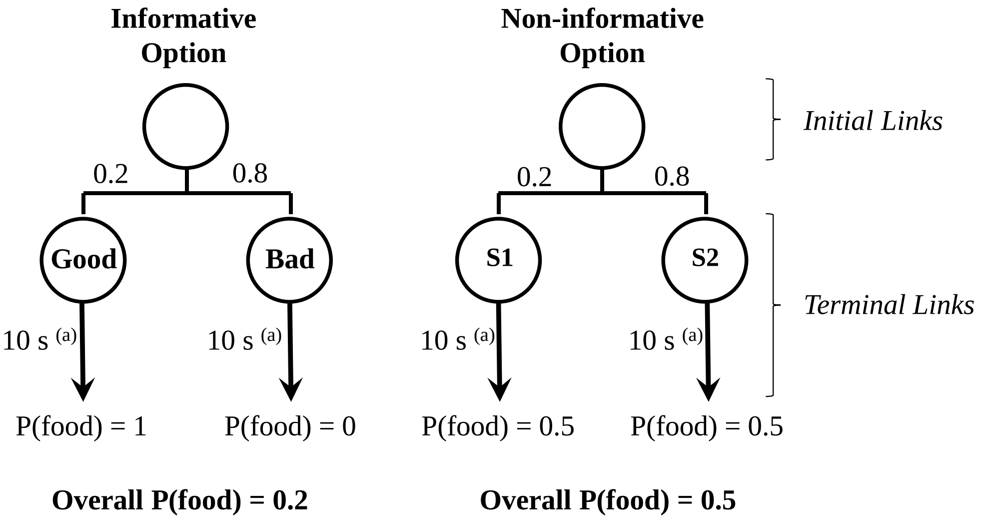


Figure 1. Schematic of outcomes for each option in the first condition of Experiments 1 and 2. In the 20-s Terminal Links Phase of Experiment 2 the terminal links (a) increased to 20 s.

Several studies with both pigeons and starlings suggest that animals choose suboptimally because the information conveyed by the terminal-link stimuli in the Informative Option is asymmetrically used: In contrast with the good-news stimulus, animals do not seem to take into account neither the probability of presentation nor the duration of the bad-news stimulus, given that they do not seem to influence the overall value of the Informative Option (Study 1; Mazur, 1991, 1995, 1996; McDevitt, Spetch, & Dunn, 1997; Pisklak, McDevitt, Dunn, & Spetch, 2015; Spetch, Mondloch, Belke, & Dunn, 1994; Stagner, Laude, & Zentall, 2012; Vasconcelos et al., 2015; Zentall, Laude, Stagner, & Smith, 2015). On the other hand, both stimuli of the Non-informative Option seem to be taken into account. The fact that animals consider all the outcomes in the Non-informative Option but not in the Informative Option leads to a distorted perception of reinforcement rates, with the Informative Option perceived as always delivering food (because only the good-news stimulus is taken into account), and the Non-informative Option perceived as delivering food only in some trials. Because of this bias, the Informative Option seems richer and is therefore preferred. It is noteworthy, however, that this misrepresentation of the overall rates of reinforcement cannot be attributed to a lack of knowledge about the contingencies associated with each terminal-link stimulus (see Figure 1). In fact, animals learn the outcomes associated with each signal: For example, when given a choice between different pairs of terminal stimuli, starlings almost never chose the bad-news stimulus, and almost always chose the stimulus associated with the higher probability of reinforcement (Vasconcelos et al., 2015; see also Stagner et al., 2012 with pigeons).

According to Vasconcelos and colleagues (2015), the decision mechanism implemented in this experimental task evolved such that animals should pay attention and engage with stimuli associated with food and ignore or even avoid stimuli associated with the absence of food. More specifically, if a stimulus associated with the absence of food is sighted, the animal uses this information to change its search pattern: It moves away from that stimulus and searches for another food source (Stephens, Brown, & Ydenberg, 2007; Stephens & Krebs, 1986). Because the animal moves away from the stimulus, the properties of that stimulus (such as its duration) are ignored, and therefore it does not affect forthcoming decisions. When, in a laboratory setting, animals are forced to stay in the presence of a bad-news stimulus, information about the absence of food cannot be used to change the current situation. However, since the animal's mechanism to deal with this situation is the one evolved in the natural environment, in the laboratory they also attempt to escape from that stimulus (by disengaging from the task, for example), which in practice means that they disregard this stimulus. Ignoring the bad-news stimulus may be optimal in the natural environment because there is little to no cost associated with ignoring the stimulus, but it is suboptimal when the animal is forced to pay the cost of waiting as in laboratory situations. This mismatch between the domain of selection (the natural environment) and the domain of testing (the laboratory) of a behavior, may underlie this apparently paradoxical behavior (Houston, McNamara, & Steer, 2007; Kacelnik, 2003; Stephens & Anderson, 2001; Stephens & Dunlap, 2009; Stephens, Kerr, & Fernández-Juricic, 2004; Stephens & McLinn, 2003; Stevens & Stephens, 2010). Based on these ideas and on optimal foraging theory, Vasconcelos et al. (2015) developed the Reinforcement Rate Model (RRM) to account for results observed in tasks similar to that depicted in Figure 1.

The RRM assumes that that, in choice situations, animals prefer the option that maximizes the long-term intake rate. To compute the rate, consider a predator searching for food. After a while, it encounters prey *i* and starts pursuing it. Then, with probability p_i, the predator catches the prey and consumes it, and with probability $1 - p_i$ the prey escapes and the predator starts a new search. Assuming prey types of equal energy content (1 unit), the rate of intake with prey *i* (R_i) is given by:

$$R_i = \frac{p_i}{s + p_i \times \left(t + h\right) + (1 - p_i) \times t} \tag{1}$$

where p_i is the probability of capturing prey i, s is the average search time, t the pursuing time and h the handling time.

In the experimental situation, p_i is the probability of reinforcement in one option, s is the intertrial interval (ITI), t is the duration of the terminal-link stimuli and h is the reinforcement duration. To predict preference in the task depicted in Figure 1, Equation 1 is applied to both the Informative and Non-informative options and the option with higher reinforcement rate should be chosen. Objectively, the rate is higher in the Non-informative Option and should thus be preferred. However, this is in sharp contrast with a myriad of experimental data (i.e., animals show a strong, often exclusive, preference for Informative Option).

In order to account for the preference for the Informative Option, and to make the reinforcement rate equations ecologically more plausible, Vasconcelos et al. (2015) modified them. First, in the Informative Option they eliminated the time the animal spends in the presence of the bad-news stimulus, because, as previously explained, in natural settings animals do not take this duration into account. Second, in both the Informative and Non-informative options the duration of the ITI (s in Equation 1) was also eliminated because: (a) it is common to all outcomes (see detailed explanation in Vasconcelos et al., 2015) and (b) it occurs *after* the reward and *before* the choice response. The second reason is supported by evidence showing that the only durations affecting preference in choice tasks are those that occur *between* choice and reward (Bateson & Kacelnik, 1996; Logue, Smith, & Rachlin, 1985; Mazur, 1989; Spetch, Belke, Barnet, Dunn, & Pierce, 1990; Spetch et al., 1994; Stephens & Anderson, 2001). With these modifications, the *perceived* reinforcement rate of the Informative and Non-informative options is given by Equations 2 and 3, respectively:

$$R_{Info} = \frac{p_{Info}}{p_{Info} \times (t+h)} = \frac{1}{t+h} \tag{2}$$

$$R_{Non-\inf o} = \frac{p_{Non-\inf o}}{p_{Non-\inf o} \times (t+h) + (1 - p_{Non-\inf o}) \times t} = \frac{1}{\dfrac{t}{p_{Non-\inf o}} + h} \tag{3}$$

where p_{Info} and $p_{Non\text{-}info}$ are the overall probabilities of reinforcement in the Informative and Non-informative options, respectively.

With this model, Vasconcelos and colleagues (2015) predicted that (a) the probability of reinforcement in the Informative Option should not affect preference for this alternative, except when $p_{Info} = 0$, and; (b) if the duration of uncertainty about reward delivery is equalized across options - by implementing a delay between choice and the signals in the Informative Option - preference should reverse to the option with higher probability of reinforcement. These and other predictions were confirmed.

One of the assumptions of the Vasconcelos et al.'s (2015) model is that, in the wild, when animals see a stimulus associated with the absence of food, they use that information to improve the current situation, that is, they move away from the stimulus to search for another patch. This "escape" behavior from the bad-news stimulus improves the situation because, as the cost of waiting for a sure no-reward is not payed, the average delay to food decreases. On the other hand, when animals in the laboratory see a stimulus associated with the absence of food, they cannot use that information to improve the current situation (i.e., they must experience the programmed delay). If animals in the laboratory use the same decision mechanisms as in the wild then, when given a choice between continuing in the presence of the bad-news stimulus and escaping from it (which would decrease the average delay to food), they should always escape.

Many researchers have implied that the bad-news stimulus is ignored or not taken into account (e.g., McDevitt, Dunn, Spetch, & Ludvig, 2016; Vasconcelos et. al, 2015; Zentall, 2016). Exploring under which conditions animals escape the bad-news stimulus also allow us to better define what it means that the bad news are *ignored* or *not taken into account*. If, on the one hand, ignoring the bad-news stimulus means that the animal is subjectively in a situation where the bad-news stimulus was never presented, animals should never escape from this stimulus: there is no reason to escape something that it is not present. On the other hand, ignoring the bad-news stimulus may mean that, even though the animal is aware of its presence (for example, it times the duration of the stimulus), it simply does not associate the presentation of this stimulus with the choice of the Informative Option (i.e., it is a problem of misattribution). In this case, the animal should escape the bad-news stimulus because by doing so the delay to the next possible reinforcer decreases. The prediction of the RRM that animals should escape the bad-news stimulus presupposes this second case: animals are aware of the bad-news stimulus, but due to a misattribution, this stimulus does not influence the value of the Informative Option.

The main goal of the present experiments was to test whether pigeons will escape from the bad-news stimulus, which indirectly allows us to better define the meaning of *ignoring* the bad-news stimulus. To that end, we implemented a modified version of a sequential-encounters procedure (Lea, 1979). In general terms, when the terminal-link stimulus appeared we gave animals a choice between continuing in the situation (that is, accepting the signal and its outcome) and escaping from the situation, thus starting a new trial.

This procedure has been extensively used to simulate a foraging situation using operant conditioning (Abarca & Fantino, 1982; Abarca, Fantino, & Ito, 1985; Fantino & Preston, 1988; Freidin, Aw, & Kacelnik, 2009; Hanson & Green, 1989a, 1989b; Lea, 1979; Mazur, 2007, 2008). In order to predict the circumstances under which animals should accept or reject a prey item these studies frequently resorted to the classical prey choice model (Charnov, 1976). Even though a first approach with the RRM predicts that animals should always escape from the bad-news stimulus, the application of the prey choice model may predict otherwise. In what follows, we briefly present the prey choice model and then apply its core ideas to the RRM in order to obtain more ecologically plausible predictions.

Stated simply, the prey model predicts that animals should accept or reject an item based on which action yields a higher long-term reinforcement rate. Note, however, the difference between this long-term maximization and the long-term maximization previously described for the RRM. For the latter, we calculated the rate of *each option* separately (Informative and Non-informative options), and predicted that the one yielding the highest rate should be chosen. With the prey model, we do not compare the rate of each option because we are not interested in predicting preference between the options. Instead, we calculate the rate of *each action* (accept or reject), that is, we calculate the average rate of food intake (with the Informative and Non-informative options, together) under two scenarios: (1) animals always accept the bad-news stimulus (and all the other stimuli; a generalist strategy) and (b) animals always reject the bad-news stimulus (but accept all other stimuli; a specialist strategy). The prediction is that the action leading to the highest rate should be always implemented because intermediate proportions of rejection never lead to the maximization of food intake in these scenarios (see proof in Charnov, 1976; Stephens & Krebs, 1986). In summary, the bad-news stimulus should always be rejected if, with this course of action, the overall rate of reinforcement is higher than when this same stimulus is accepted, that is, when $R_{reject} > R_{accept}$.

To calculate R_{reject} and R_{accept} some assumptions of the RRM were modified as the situation itself was modified. More specifically, in its original instantiation, the model assumed that animals ignore both the bad-news stimulus and the ITI duration. However, since in the present task animals have to choose between accepting and rejecting the bad-news stimulus, they should pay attention to the consequences of each decision. Since the immediate consequence of *rejecting* is the start of the ITI, escaping should be influenced by the ITI duration. Similarly, the consequence of *accepting* the bad-news stimulus is its continuation, so accepting the stimulus should be affected by the stimulus duration. That is, the animals should take into account both the bad-news and the ITI duration in the trials in which they accept and reject, respectively.

When animals accept the bad-news stimulus (and all the other stimuli), they take into account the duration of all the terminal links, t. Moreover, when food is delivered (either in the Informative Option or in the Non-informative Option, which occur with probability p_{reward}), they also take into account the reinforcement duration, h. Animals adopting this generalist strategy will experience the long-term reinforcement rate given by

$$R_{accept} = \frac{p_{reward}}{t + p_{reward} \times h} = \frac{1}{\dfrac{t}{p_{reward}} + h} \tag{4}$$

where p_{reward} is the overall probability of reward given by the weighted average of probabilities on the Informative and Non-informative options.

When animals reject the bad-news stimulus (but accept all other stimuli), the durations taken into account depend on which stimulus is presented. When the bad-news stimulus is shown it is rejected, so the ITI duration, s, is taken into account; in the trials with the remaining stimuli, the stimulus shown is accepted and the terminal-link duration, t, is taken into account. Finally, when the trial ends with food, the reinforcement duration, h, is also taken into account as in the original RRM instantiation. Animals adopting this specialist strategy will experience the long-term reinforcement rate given by

$$R_{reject} = \frac{p_{reward}}{p_{bad-news} \times s + \left(1 - p_{bad-news}\right) \times t + p_{reward} \times h} \tag{5}$$

From Equations 4 and 5 we predict that animals should escape from bad news when

$$R_{reject} > R_{accept}$$

$$\frac{p_{reward}}{p_{bad-news} \times s + \left(1 - p_{bad-news}\right) \times t + p_{reward} \times h} > \frac{1}{\dfrac{t}{p_{reward}} + h}$$

which simplifies to

$$s < t \tag{6}$$

Equation 6 shows that animals should escape the bad-news stimulus if the ITI duration is shorter than the terminal-link durations. Neither the probability of reinforcement (p_{reward}) nor the reward duration (h) should influence the decision to escape. Figure 2 shows the reinforcement rate, both when accepting (R_{accept}, white dots, Equation 4) and escaping (R_{reject}, black dots, Equation 5) the bad-news stimulus. The left panel shows the rates as a function of the ITI duration, with the terminal-link durations constant at 10 s; the right panel shows the rates as a function of the terminal-link duration, with the ITI duration constant at 10 s.

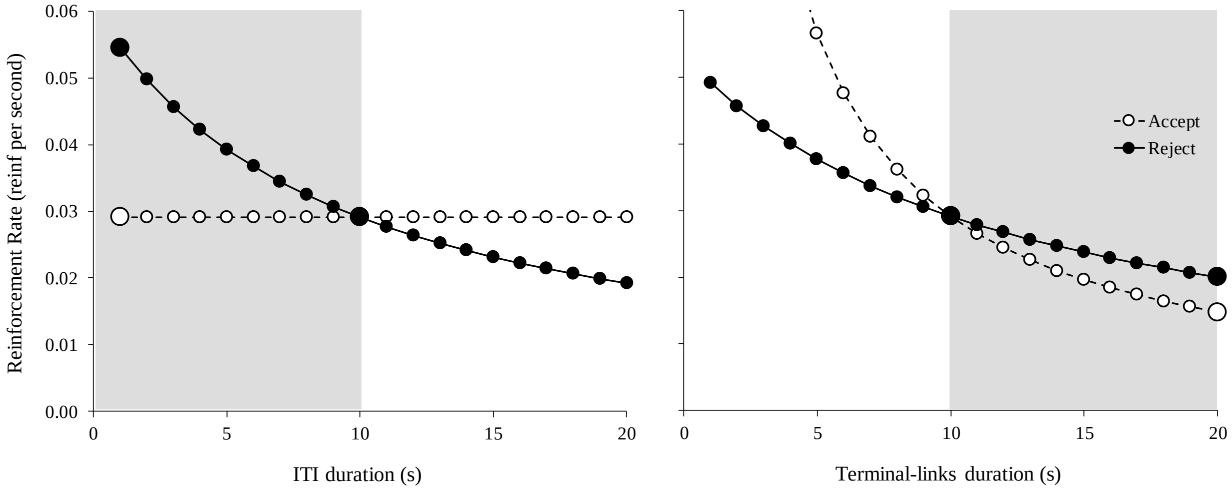


Figure 2. Reinforcement rate functions when accepting (white dots) and rejecting (black dots) the bad-news stimulus, as a function of ITI duration (left panel) and terminal-link duration (right panel). Rejecting the bad-news is the best action within the shaded region. The larger dots identify the values tested in the present experiments.

Since R_{accept} is independent of the ITI duration (cf. Equation 4), its value remains constant (Figure 2, left panel). On the other hand, R_{reject} is strongly influenced by the ITI duration because when the bad-news stimulus is rejected the animal takes this duration into account. Therefore, as the ITI increases the R_{reject} decreases. It is predicted that when the ITI is shorter than 10 s (the duration of the terminal links), animals should reject the bad-news stimulus ($R_{reject} > R_{accept}$; shaded area), and when it is longer than this value, they should accept the stimulus ($R_{accept} > R_{reject}$; white area).

While the ITI duration affects only the value of R_{reject}, the terminal-link duration influences both R_{reject} and R_{accept}. Specifically, as the duration of the terminal-link signals increase, both reinforcement rates decrease (Figure 2, right panel). In this case, when the duration of the terminal links is longer than 10 s (the duration of the ITI), animals should reject the bad-news stimulus ($R_{reject} > R_{accept}$; shaded area) and when it is shorter than that value, they should accept it ($R_{accept} > R_{reject}$; white area).

The main goal of the present experiments was to test the prediction of Vasconcelos et al.'s (2015) model that animals should escape the bad-news stimulus. A second goal was to test if escaping was affected by either the ITI or the terminal-link durations. Even though optimality can only be achieved by either always or never escaping, several experiments have shown that the probability of escaping is usually intermediate (Abarca & Fantino, 1982; Abarca, Fantino, & Ito, 1985; Fantino & Preston, 1988; Freidin et al., 2009; Hanson & Green, 1989a, 1989b; Lea, 1979; Mazur, 2007, 2008). As a consequence, we predicted that the probability of escaping was dependent on the relative gain from escaping vs. accepting the bad-news stimulus. For example, shortening the ITI duration should lead to more escape responses compared with increasing the terminal-link duration (compare the distance between the black and white dots in Figure 2). In Experiment 1, we manipulated the ITI duration and in Experiment 2 we manipulated the terminal-link duration.

2. Experiment 1 – Effect of the ITI duration in escaping "bad news"

2.1. Method

2.1.1. Subjects

Eight pigeons (*Columba livia*) with previous experience in timing procedures participated. The birds were housed individually in a room with controlled temperature (around 21 ºC), and with a 11h:13h dark/light cycle (lights on at 8 am). In their home cages, animals had water and grit always available. The birds started the experiment when they were at 80% of their free-feeding weights.

2.1.2. Apparatus

Four operant chambers for pigeons were used. All chambers were enclosed in an opaque box equipped with a fan that circulated air and masked outside noises. Three of the chambers were manufactured by Med Associates, and were 28.5-cm high, 24-cm long and 30-cm wide. The front panel had three horizontally-aligned keys, each 2.5 cm in diameter, 6 cm apart, center-to-center. The bottom edge of the keys was 21 cm from the floor. Below the keys there was a 6-cm wide x 5-cm high feeder opening, centrally aligned and 4 cm above the floor. When activated, food was illuminated with a 1.1-W light. On the panel opposite to the front panel, 23 cm above the floor, there was a 2.8-W light that illuminated the whole box. The fourth chamber was custom-built and it was equal to the other boxes except in some measures. This box was 31-cm high, 33-cm long and 33-cm wide. The keys were 9 cm apart, center-to-center. The feeder opening was 6-cm wide x 4.5-cm high, 6.5 cm above the floor. The houselight was 27.5 cm above the floor.

A personal computer with a custom Visual Basic 2010 program controlled the events and recorded data via Whisker (Cardinal & Aitken, 2010).

2.1.3. Procedure

Pretraining. Pigeons received one session in which they had to peck each of the stimuli used in this task: red, green, yellow, blue and white on the left and right keys, and white and a vertical white bar on a black background on the center key. Each session had 48 trials, 4 trials per stimulus. Each trial began with the illumination of one of the cues, and one peck at the illuminated key was followed by a 3-s reinforcer. The trials were separated with a 10-s ITI with only the houselight on. The houselight turned off when the trial started. The peck requirement was increased to 10 pecks during the second and third pretraining sessions.

Terminal-Links Escape. Each session was composed of 120 trials, divided into four 30-trial blocks. Within each block, there were 20 forced- (10 per option) and 10 free-choice randomly intermixed trials. In a forced-choice trial only one option was presented.

At the beginning of a forced-choice trial the center key started flashing a white hue (250 ms on, 250 ms off). A peck at this key turned it off and turned on one of the side keys with a white hue (initial link, Figure 1). For half of the birds the Informative Option was always on the left and the Non-informative Option always on the right; for the remaining birds the opposite was true. A peck on the white side key initiated the 10-s terminal link. If the Informative Option was presented, on 20% of the trials the good-news stimulus was turned on (e.g., red), and after 10 s food was delivered; on the remaining 80% of the trials the bad-news stimulus was turned on (e.g., green), and 10 s after the trial ended without reinforcement. If the Non-informative Option was presented instead, on 20% of the trials one stimulus was turned on (e.g., yellow) and on the other 80% of the trials a different stimulus (e.g., blue) was turned on. In both cases, the stimulus stayed on for 10 s and was followed by food on half of the trials. All trials ended with a 10-s ITI, during which only the houselight was on. The terminal-link colors associated with each option were counterbalanced across pigeons, with the constraint that red and green were always associated with the one option and yellow and blue were always associated with the other option.

The free-choice trials were exactly the same as the forced-choice trials, with the exception that during the initial link the two side keys were illuminated; one peck at one of the side keys turned the other key off. Afterwards, the trial proceeded as described for forced-choice trials.

Importantly, the animals could escape the trials once the terminal link was presented. In all trials, a vertical white bar turned on at the center key (escape key) simultaneously with the terminal-link stimulus. If the pigeon pecked the escape key once, the trial ended immediately without reinforcement, a 10-s ITI followed, and a new trial started. This phase lasted 10 sessions.

Forced-Escape Trials. This phase was equal to the previous one, but 40 new trials were added in which the animals were forced to peck the escape key. In these new trials, after the initial-link peck, the side key was turned off and only the vertical bar on the center key was presented (the escape key). A peck at this key turned it off and was followed by a 10-s ITI and a new trial. There were 20 forced-escape trials after an initial link on the left and 20 after an initial link on the right. These trials were equally distributed across the four blocks of 40 trials, totaling 160 trials. This phase lasted for 15 sessions.

1-s ITI. In this phase everything remained the same, with the exception that all ITIs were 1-s long. This phase lasted for 10 sessions.

Initial-Links Escape. According to Vasconcelos et al. (2015), animals should not escape in the initial links because these stimuli provide no information about impending food. This phase was the same as the previous one, with the following exceptions: (a) animals could escape during the initial and the terminal links, and (b) 40 new trials were added to force the animals to peck the escape key during the initial link.

In all forced- and free-trials, when the side(s) key(s) was/were turned on white, the vertical white bar was also presented at the center key (escape key). In order to avoid escape responses caused by the ballistic nature of the pecking patterns shown by some pigeons (i.e., they may show a burst of responses to the flashing attention key) a self-initiated fixed-interval schedule was in effect in the escape key: the first peck on this key initiated a 1-s period. After that time, if the center key was pecked again the trial was cancelled immediately, and a 1-s ITI followed. This schedule was used in order to avoid initial-link escapes caused by uninterrupted pecking to the center key after trial initiation. If they pecked one of the side keys the terminal link started as usual. In the additional 40 trials, after a peck at the center key to start the trial, only the center vertical bar was presented, with the same schedule described above. Each 50-trial block had 20 trials in which the animals were forced to escape (10 in the initial link and 10 in the terminal link), and 30 trials in which the animals could escape (10 forced-choice trials with each option and 10 free-choice trials). The sessions were composed of four blocks, totaling 200 trials. This phase lasted for 10 sessions.

2.2. Results and Discussion

Unless stated otherwise, for all analyses the mean values refer to the last three sessions of each condition. A significance level of 0.05 was used in all analyses.

Preference. Figure 3 shows the mean proportion of choices of the Informative Option over sessions. All animals showed an initial tendency to choose the optimal, Non-informative Option, a result commonly seen in the literature (e.g., Laude, Stagner, & Zentall, 2014; Zentall & Stagner, 2011). Then, all birds developed a strong preference for the Informative Option, which was significantly above chance by the fifth session, $t(7) = 3.66$, $p < .01$, d = 1.29, 95% CI for d [0.13, 2.46]. Preference for the Informative Option remained almost exclusive throughout the following phases.

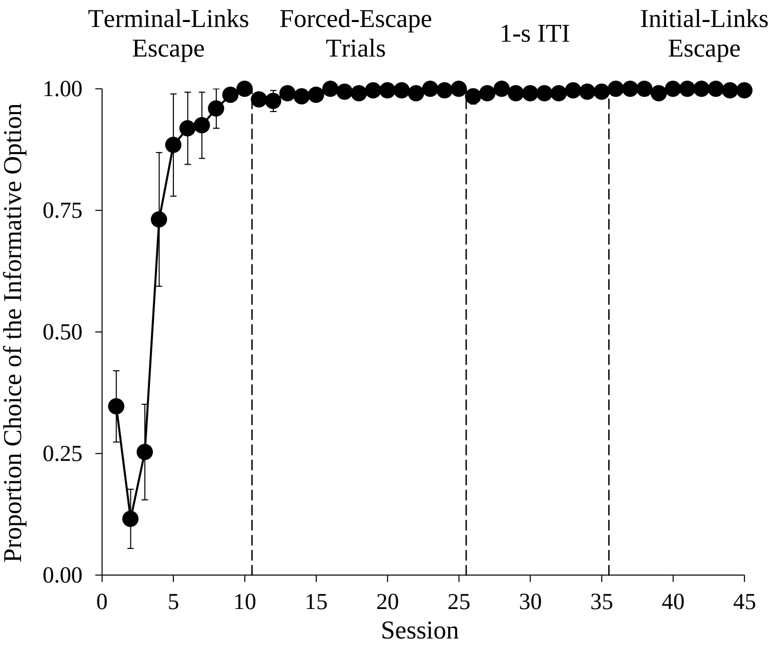


Figure 3. Mean proportion of choices of the Informative Option across sessions in Experiment 1. The error bars are the standard error of the mean.

Escape Responses. Figure 4 shows the proportion of escape responses when each of the terminal-link stimuli were on (cf. different lines). In the first session, birds escaped from all terminal-link stimuli: $M_{\text{bad news}}$ = .62 (SEM = .09), $M_{\text{good news}}$ = .28 (SEM = .11), M_{S1} = .43 (SEM = .12), M_{S2} = .48 (SEM = .11). In subsequent sessions, escaping mostly occurred in the presence of the bad-news stimulus.

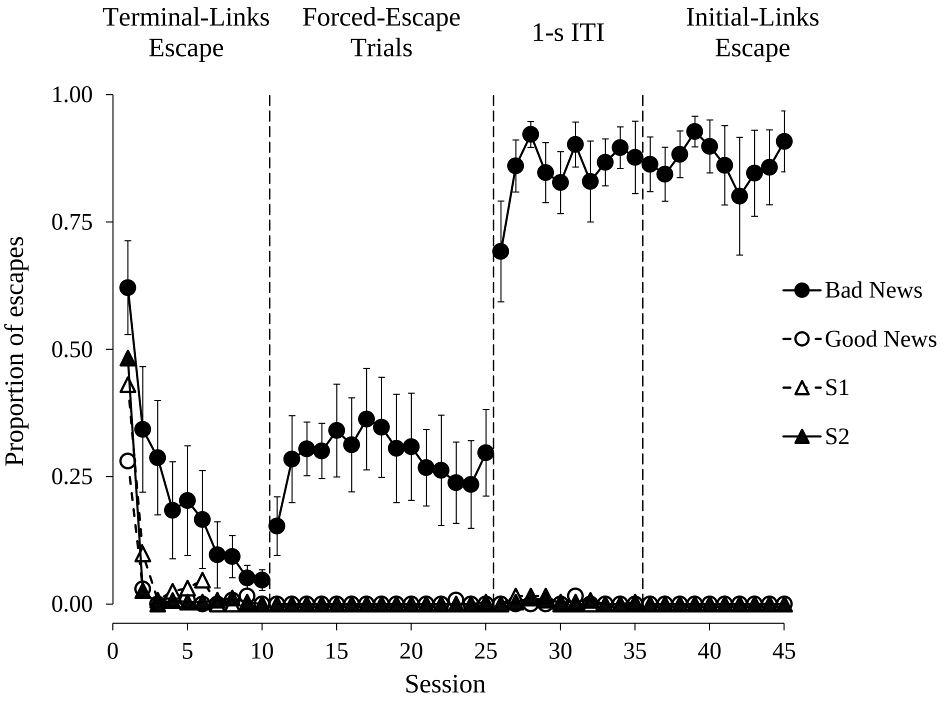


Figure 4. Mean proportion of escape responses for each terminal-link stimulus in Experiment 1. The error bars are the standard error of the mean.

During the Terminal-Links Escape Phase, escaping the bad-news stimulus decreased over sessions, with an average of 5% of escapes on the tenth session. When forced-escape trials were introduced in the Forced-Escape Trials Phase, the proportion of escapes increased and stabilized around .26 (SEM = .003), but remained below 50%, $t(7) = 3.24$, $p = .01$, d = 1.15, 95% CI [0.00, 2.29]. Finally, when the ITI was decreased to 1 s (in the last two phases), escaping increased considerably, with the average escape always above .80 (excluding the first session). The increase in the proportion of escape responses from the Forced-Escape Trials Phase to the 1-s ITI Phase was statistically significant, $t(7) = 8.90$, $p < .001$, d = 3.15, 95% CI [1.52, 4.77]. Averaged across the last three sessions, the proportion of escapes from the bad-news stimulus was significantly above chance both in the 1-s ITI Phase, $t(7) = 8.11$, $p < .001$, d = 2.87, 95% CI [1.32, 4.41] and in the Initial-Links Escape Phase, $t(7) = 5.23$, $p = .001$, d = 1.85, 95% CI [0.57, 3.13].

On the last phase, animals could escape both during the initial and terminal links, but only three out of eight birds pecked the escape key during the initial links, always in forced-choice trials with the Non-informative Option. These pecks were, however, rare occurring on 2.5% of the Non-informative trials.

In the Forced-Escape Trials Phase, the perceived reinforcement rate was the same if animals escaped or accepted the bad-news stimulus, so we expected a proportion of escapes around 50 % (cf. left panel of Figure 2). However, the proportion was below indifference, which shows a bias against escaping. This bias may have occurred because accepting was the "default" response: While escaping required a peck on a key, accepting did not require any specific behavior. This is consistent with other results: For example, in Lea's (1979) experiments, the default was escaping, that is, if pigeons did not peck any key for around 4 s, the option was rejected, but if they pecked the key the option was accepted. In a condition where 50% escaping was expected, animals consistently showed a bias towards escaping, possibly because of this difference in the requirement for escaping and accepting. Moreover, in a comparable situation, Mazur (2007) required one peck both to escape and to accept the option, and there was no bias: the proportion of escape responses was close to 50 %.

Finally, when the ITI was decreased to 1 s, the reinforcement rate was higher if pigeons escaped the bad-news stimulus. Consistent with our predictions, animals dramatically increased the proportion of escape responses in these phases

3. Experiment 2 – Effect of the terminal-link duration in escaping "bad news"

In Experiment 1, in the Forced-Escape Trials Phase, the reinforcement rate was the same regardless of whether animals accepted or rejected the bad-news stimulus. Even though pigeons used the escape key in this phase, they seemed to have a bias against escaping, which could have been the result of the different requirements for escaping and accepting the stimulus (peck vs. not peck, respectively). An alternative hypothesis is that the low frequency of escape responses was induced by avoidance of the escape key, given its spatial contiguity to the bad-news stimulus. In fact, unsystematic observations revealed that when the pigeons saw the bad-news stimulus they turned their back to it, and consequently to the escape key also. Thus, there could have been some approach-avoidance behavioral competition. In Experiment 2, we (a) placed the escape key away from the bad-news stimulus, to test whether pigeons would escape more often, and; (b) manipulated the duration of the terminal links, while keeping the ITI constant

3.1. Method

3.1.1. Subjects

Four pigeons (*Columba livia*) participated. The animals were kept in the same conditions as in Experiment 1.

3.1.2. Apparatus

One custom-built long operant box was used. The box was 33-cm high, 96-cm long, and 31-cm wide (see Figure 5). The box was placed in a shelf covered by wood panels on all sides except in the front, where it was covered by a blackout roller blind. Behind the box, two speakers, one on each side, were connected to a mini laptop that produced continuous white noise during the session.

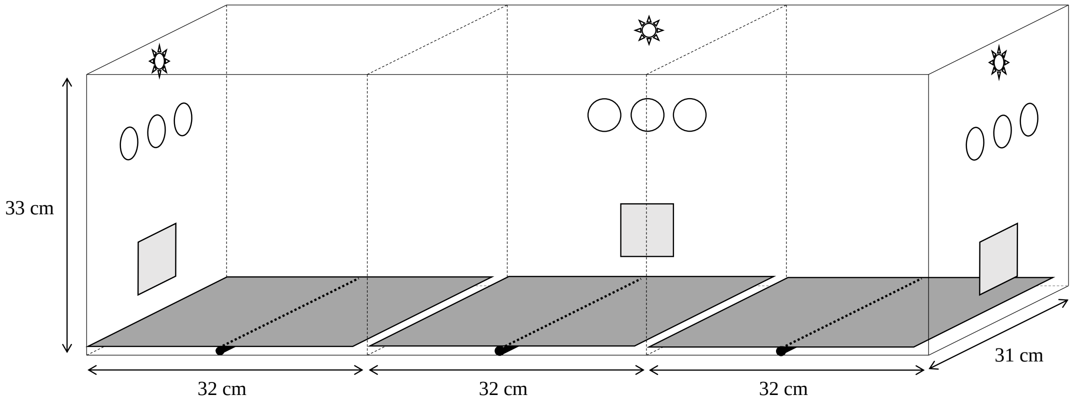


Figure 5. Diagram of the long operant chamber used in Experiment 2 (courtesy of Álvaro Viúdez).

There were three response panels in the box: one on the left side, one on the right side and one centered in the back panel. Each panel was equipped with three horizontally aligned keys, each 2.5-cm in diameter. The keys were 8 cm apart, center-to-center, and 24 cm above the floor (measure from their lowest point). A feeder opening (5-cm high x 6-cm long) was centrally located on each panel, 9 cm above the floor. When activated, the food hopper was illuminated with a 1.1-W light. Above the keys, centrally located and 1.5 cm below the ceiling, a 2.8-W houselight illuminated the box. We only used the right key of the left panel, the center key of the central panel and the left key of the right panel, henceforth referred to as the left, center and right keys, respectively.

The floor of the box consisted of three movable panels of equal size. Each panel rested on a pivot that was centrally located underneath the panel. The panel could tilt to the right or to the left, depending on where the pigeon was stepping. Below each panel, and on each side of the pivot there was a switch that registered where the pigeon was standing. There were thus 6 locations, numbered from 1 (the left side of the leftmost floor panel) to 6 (the right side of the rightmost floor panel).

A personal computer with a custom Visual Basic 2010 program controlled the events and recorded data via Whisker (Cardinal & Aitken, 2010).

3.1.3. Procedure

The procedure was similar to that used in Experiment 1, with the following exception: we first replicated the original procedure without the escape key, and only after we introduced the escape key. The replication was an important step given that this task has never been implemented in a long operant box.

Pretraining. None of the birds had experience with a long operant box, so they were first trained to eat from each feeder. Then, each bird received two sessions with 44 trials in which one peck at an illuminated key - red, green, yellow, blue and white on the left and right keys, and white on the center key - was required to get a 3-s access to food. The reinforcer was delivered on the feeder located on the same panel as the pecked key. All three houselights were always on except during reinforcement, and a 10-s ITI separated the trials. The peck requirement was then increased to 10 pecks for 2 additional sessions.

Training. Each session consisted of four 30-trial blocks, totaling 120 trials. Each block had 10 forced-choice trials for each option and 10 free-choice trials, all randomly intermixed.

A forced-choice trial started with the center key flashing with a white hue (250 ms on, 250 ms off). A single peck to this key turned it off and one of the white side keys turned on (initial link, Figure 1). The Informative and Non-informative Options were always on the same side for each bird, but the positions were counterbalanced across birds. If the Informative Option was on, a single peck turned the initial link off and the good-news stimulus was presented on 20% of the trials. This stimulus stayed on for 10 s and was always followed by food; on the remaining 80% of the trials, the bad-news stimulus was presented for 10 s and the trial ended without food. If the Non-informative Option was presented instead, a single peck turned the initial link off and one of two stimuli (S1 or S2) was presented for 10 s (with probabilities 20% and 80% as in Figure 1), and was followed by food on 50% of the trials. During the whole session, except during reinforcement, the three houselights were on. The trials ended with a 10-s ITI.

In the free-choice trials the two side keys were illuminated during the initial link. When the animal chose one option by pecking it, the other key turned off and the terminal link in the chosen key started. This phase lasted for 10 sessions.

Reversal. In this phase, the location of the Informative and Non-informative options was reversed. This phase ensured that the results observed in the previous phase reflected a preference for one of the options and not a preference for one of the side keys. In all other respects the sessions remained the same as in the previous phase. This phase also lasted for 10 sessions.

Terminal-Links Escape. In this phase birds could escape the terminal links in all trials. After pecking the white hue in one of the side panels, the white central panel key (the escape key) turned on. If the pigeon moved to the central panel and pecked the escape key during the terminal link, all keylights were turned off and a 10-s ITI followed.

Forty trials in which the animals were forced to peck the escape key were also added. In a forced-escape trial, after one of the side keys was turned on white (initial link), one peck at that key turned it off and the central panel escape key turned on. The trial only proceeded when the pigeon pecked the escape key, which was followed by the ITI. Note that, even though the white central panel key was used both to start the trial and to escape the trial, its function was visually signaled: When it was to start the trial the keylight was flashing and when it was to escape the trial it was steady. The sessions were divided into four 40-trial blocks (10 free-choice trials, 10 forced-choice trials of each option and 5 forced-escape trials of each option), totaling 160 trials. This phase lasted for 10 sessions.

20-s Terminal Links. The sessions were as in the previous phase with two exceptions. First, all terminal links were increased from 10 s to 20 s. Second, sessions were composed of two 40-trial blocks, totaling 80 trials. This phase lasted for 20 sessions. The data of each two sessions were collapsed in order to make it more comparable to the previous phase.

3.2. Results and Discussion

Unless stated otherwise, for all analyses the mean values refer to the last three sessions of each condition. A significance level of 0.05 was used in all analyses.

Preference. The mean proportion of choices of the Informative Option over sessions is plotted in Figure 6.

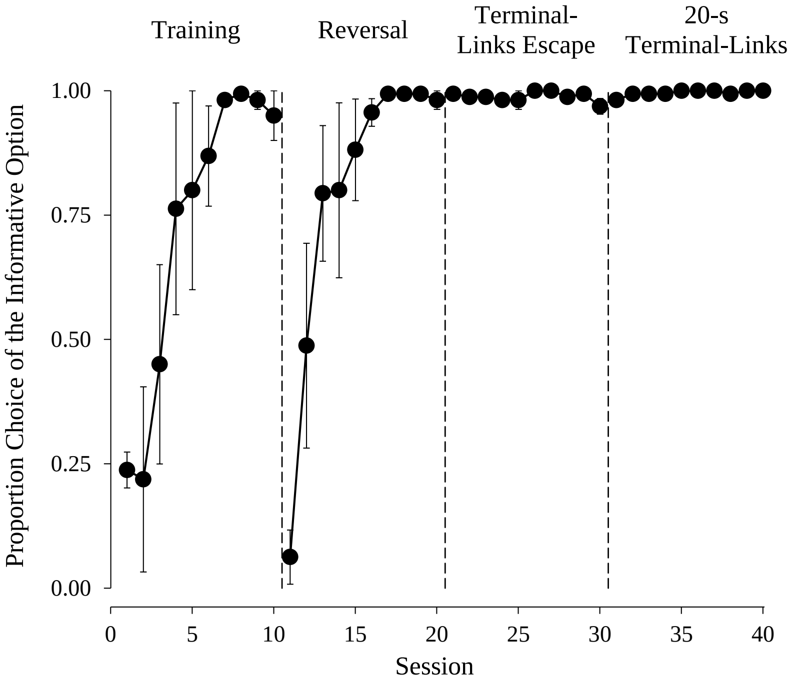


Figure 6. Mean proportion of choices of the Informative Option across sessions in Experiment 2. The error bars are the standard error of the mean.

During the Training Phase, all birds acquired a strong preference for the Informative Option (M = .98, SEM = .03) that was significantly above chance, *t*(3) = 19.00, *p* < .01, d = 9.50, 95% CI [3.09, 15.91]. In the first session of the Reversal Phase, all animals preferred the Non-informative Option, M = .06, SEM = .05 (i.e., the previously preferred side panel). After a few sessions, all birds reversed their choice, showing a preference for the Informative Option significantly above chance (M = .99, SEM = .01), *t*(3) = 62.07, *p* < .01, d = 31.04, 95% CI [10.70, 51.38]. This preference remained stable and high (> 93%) in all subsequent phases.

Escape Responses. The birds never escaped during the good-news stimulus or during the terminal links of the Non-informative Option. However, they did escape from the bad-news stimulus (Figure 7).

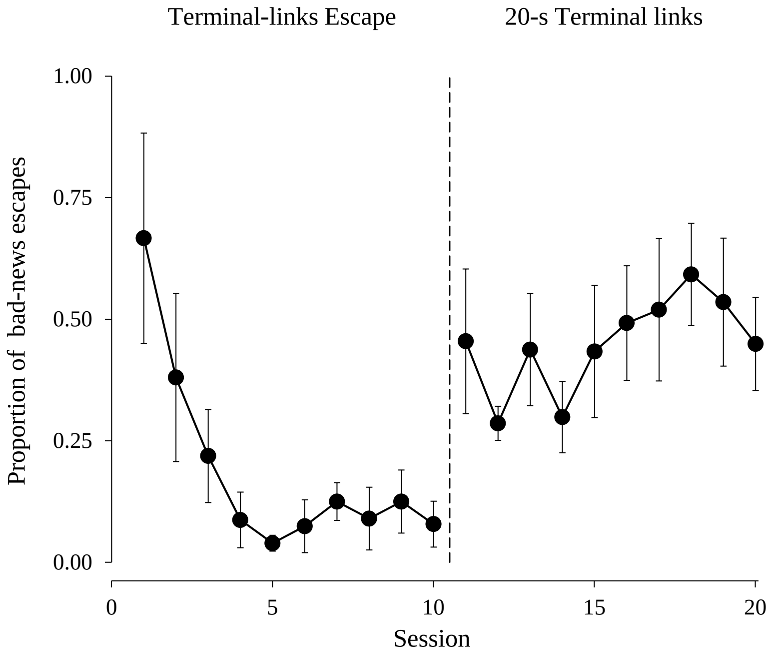


Figure 7. Mean proportion of escape responses during the bad-news terminal link. The error bars are the standard error of the mean.

Figure 7 shows the proportion of escape responses when the bad-news stimulus was presented. Similarly to Experiment 1, when the escape key was first introduced – Terminal-Links Escape Phase - animals pecked at it, but escaping decreased over sessions. The proportion of escapes during the Terminal-Links Escape Phase stabilized around .10 (SEM = .06), and remained below chance levels, *t*(3) = 7.24, *p* < .01, d = 3.62, 95% CI [0.76, 6.47]. When the terminal links increased from 10 to 20 s, escape proportion increased to .53 (SEM = .11), that is, to a chance level, *t*(3) = 0.24, *p* = .83, d = 0.12, 95% CI [-1.48, 1.72]. The increase in the proportion of escapes from the Terminal-Links Escape Phase to the 20-s Terminal links Phase was statistically significant, *t*(3) = 5.31, *p* = .01, d = 2.65, 95% CI [0.29, 5.01].

When both the terminal links and the ITI were 10-s long, there was no advantage of escaping over accepting the bad-news stimulus in terms of the perceived overall reinforcement rate (cf. right panel of Figure 2). However, as in Experiment 1, pigeons showed a bias against escaping, which strengthens the idea that this occurred because escaping required a peck, whereas accepting did not. Also, we hypothesized that in Experiment 1 pigeons did not peck the escape key more often because there was a tendency for animals to move away from the bad-news stimulus, which was next to the escape key, thus precluding the animals from pecking the escape key. In this experiment, the distance between the escape and the bad-news keys dramatically increased, and in phases with the same parameters (Forced-Escape Trials Phase in Experiment 1 and Terminal-Links Escape Phase in Experiment 2), the proportion of escapes was not higher in Experiment 2. In fact, escaping decreased from .26 in Experiment 1 to .10 in Experiment 2, which leads to the conclusion that the proportion of escapes in Experiment 1 was not fully explained by the relative position of the escape key. As discussed before, the bias against escaping was most likely due to the different peck requirements of accepting and rejecting the bad-news stimulus.

Movement. The movement of the birds during the terminal links was registered by the activation of the floor switches. Figure 8 shows the average position of the birds in the trials in which they had no means of escaping (Training and Reversal phases) or did not escape the terminal link (Terminal-Links Escape and 20-s Terminal Links phases). Each line represents a terminal-link stimulus. During the Training Phase, time into the terminal link was divided in 1-s bins, and in subsequent phases 0.5-s bins were used. To simplify, data is presented as if for all birds the Informative Option was on the left (box location = 1) and the Non-informative Option on the right (box location = 6) during the Training Phase.

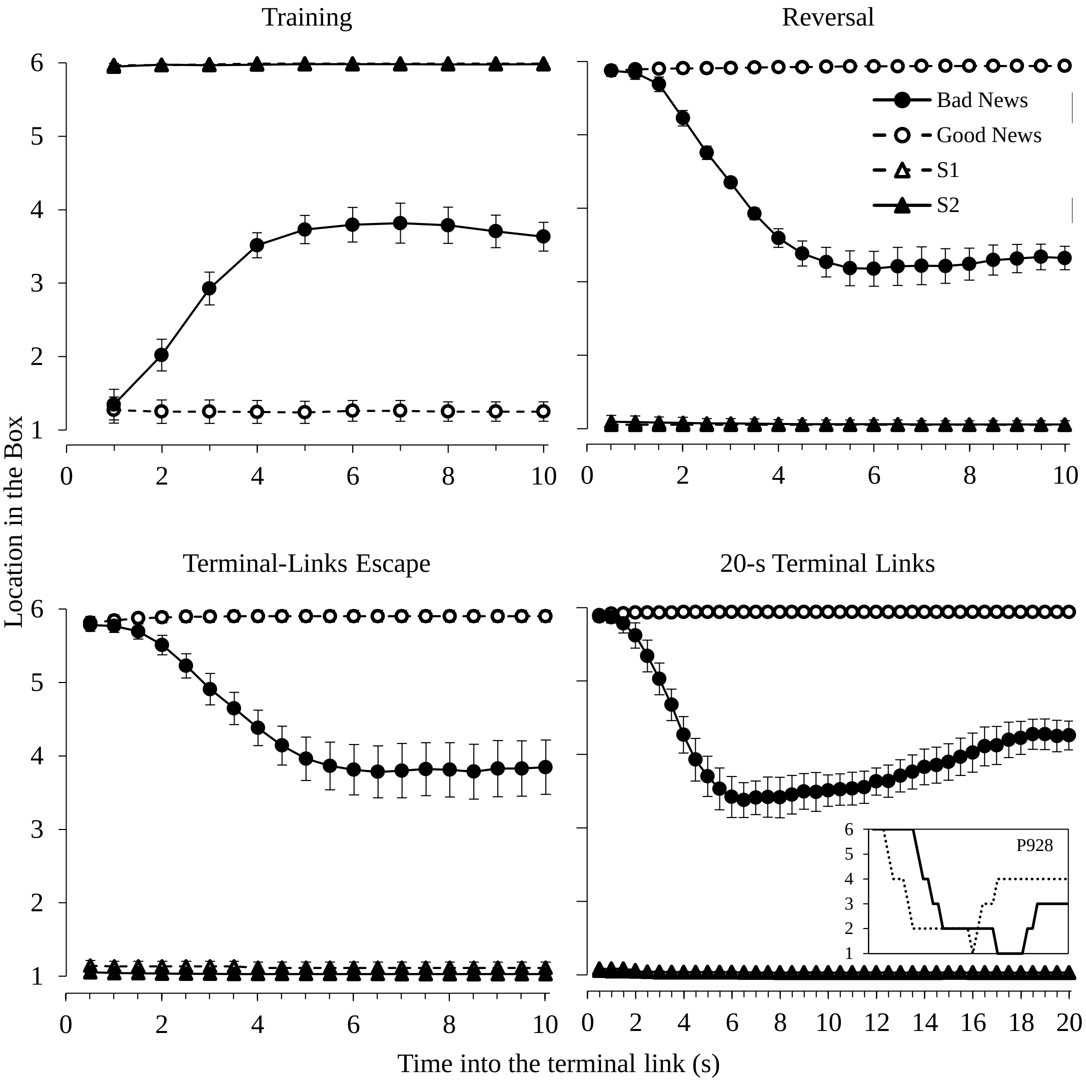


Figure 8. Mean location in the long operant box during each terminal-link stimulus. Each panel shows the results of a phase of Experiment 2. The error bars are the standard error of the mean. The inset panel (bottom right panel) shows two bad-news trials for pigeon P928.

The top left panel of Figure 8 shows that during the Training Phase pigeons stayed near the chosen key except when the bad-news stimulus was presented. In this case, pigeons moved to the center of the box and waited to start the next trial.

After the first phase, the position of the keys was reversed, but the animals adapted to the change, exhibiting the same motion pattern: They stayed near the good-news, S1 and S2 stimuli and moved away from the bad-news stimulus (Figure 8, top right and bottom panels). During the 20-s Terminal Links Phase (Figure 8, bottom right panel), the motion pattern was slightly different: When animals saw the bad-news stimulus they moved towards the side of the Non-informative Option (locations 1 and 2) and then moved to the center of the box (locations 3 and 4). To illustrate, the inset panel shows the movement during two bad-news trials for one pigeon. Because the time at which animals arrived near and departed from the Non-informative Option side varied from trial to trial, the average function seen in Figure 8 is smoother than the actual movement during a single trial. Moving further away from the negative stimulus could be either the result of greater aversiveness of the bad-news stimulus when it was presented for 20 s, or simply the result of having more time to move.

Even though the motion pattern observed when animals did escape the bad-news stimulus is not presented, it was similar to that shown in Figure 8, with the exception that when animals arrived to the center of the box they pecked the escape key, so the total duration of the terminal link was shorter than when animals did not escape.

These results confirm what we had previously observed in regular operant chambers: When animals face the bad-news stimulus, they move away from it (this is also consistent with Dinsmoor, Mueller, Martin, & Bowe, 1982; Rand, 1977). When the escape key was introduced, even though pigeons did not always escape, they did move away from the bad-news stimulus in most of the trials. In particular, they moved to the center of the box (floor panels 3 and 4) or the side opposite to the bad-news key (floor panels 1 and 2) in 93% (SEM = 3%) of the trials in the Terminal-Links Escape Phase and in 98% (SEM = 1%) of the trials in the 20-s Terminal links Phase. This confirms that animals nearly always physically escaped from the bad-news stimulus, even though they did not always start the next trial by pecking the escape key.

4. General Discussion

When animals are given a choice between 20% signaled reinforcement and 50% unsignaled reinforcement, they consistently choose the option that yields less reinforcement but immediately informs whether food will be delivered (e.g., Stagner & Zentall, 2010, Zentall & Stagner, 2011, Stagner et al., 2012; see Zentall, 2016 and McDevitt et al., 2016 for a review). The most consistent hypothesis to explain all the results with this task is that animals take into account the stimuli that are associated with food – both in the Informative and Non-informative Options – but ignore the stimulus that is never followed by food – in the Informative Option. This selective attention and engagement with different stimuli leads pigeons to disregard the cost of waiting in the Informative Option but not in the Non-informative Option, resulting in a higher perceived reinforcement rate in the Informative

Option. In the present experiments we further explored what it means to *ignore* the bad-news stimulus.

Vasconcelos and colleagues (2015) asserted that the suboptimal choice in this task results “from testing the animals in a situation where information is useless, while the birds’ psychological processes are adapted to a world in which information alters the subsequent behavior” (p. 9). Thus, in the present experiments we made information useful, by allowing birds to escape from the stimulus shown, and we expected pigeons to use that information in an appropriate manner. To make specific predictions, we adapted both the Vasconcelos et al.’s (2015) model and the prey model (Charnov, 1976). We predicted that: (a) animals should only escape from the bad-news stimulus; (b) escaping should increase as the ITI decreases; (c) escaping should increase as the terminal-link duration increases, and; (d) the probability of escaping should be associated with the degree of improvement in the reinforcement rate when escaping vs. accepting the bad-news stimulus.

Regarding (a), in both Experiments 1 and 2 we showed that, when animals escape, they only escape in the presence of the bad-news stimulus (they do not escape in the presence of other terminal-link stimuli, nor during the initial links). Moreover, consistent with other studies with the successive-encounters procedure, escaping was not all-or-none, but assumed intermediate values (Abarca & Fantino, 1982; Abarca, Fantino, & Ito, 1985; Fantino & Preston, 1988; Freidin et al., 2009; Hanson & Green, 1989a, 1989b; Lea, 1979; Mazur, 2007, 2008).

The fact that animals escape the bad-news stimulus indicates that this stimulus was not ignored, that is, animals did not behave as if the stimulus was not present. Escaping from the bad-news stimulus entails an identification of this stimulus as a signal for the absence of food. Moreover, in all conditions animals showed an almost exclusive preference for the Informative Option, which indicates that, even though the choice of this option was followed by the bad-news stimulus 80% of the time, this stimulus may not have been attributed to this option. For example, when animals only escape from the bad-news stimulus in about 26% of the trials (in the Terminal-Links Escape Phase of Experiment 1; Figure 4), they also showed an exclusive preference for the Informative Option (Figure 3). If the introduction of the escape key made animals pay attention to the bad-news stimulus and to associate this stimulus with the Informative Option, then the 74% of the trials in which birds did not escape the bad-news stimulus should have affected the value of the Informative Option. Nonetheless, preference for this option remained high instead of decreasing. Together, the strong preference for the Informative Option and the escape responses from the bad-news stimulus indicate that this stimulus was *ignored* or *not taken into account* in the sense that the animal did not associate it with the choice of the Informative Option. This in turn explains the staggering lack of effect of increased the bad-news stimulus duration to 200 s on the value of the Informative Option reported in Study 1. As the previous explanations for this suboptimal choice do not explicit what is the nature of *not taking into account* the bad-news stimulus, the results and implications of the present experiments are crucial for a better understanding of the choice mechanism involved in this task.

Regarding (b) and (c), in Experiment 1 we showed that decreasing the ITI duration from 10 to 1 s significantly increased the proportion of escape responses from the bad-news stimulus and, in Experiment 2, we showed that increasing the terminal-link duration from 10 to 20 s also led to increases in the proportion of escape responses. Regarding (d), and to compare the results of the two experiments, Figure 9 shows the observed proportion of bad-news escape responses in each phase (black dots) and the proportion of increase in the reinforcement rate when escaping rather than accepting the bad news stimulus (white dots). The latter was given by $(R_{\text{Escape}} - R_{\text{Not Escape}})/(R_{\text{Not Escape}})$.

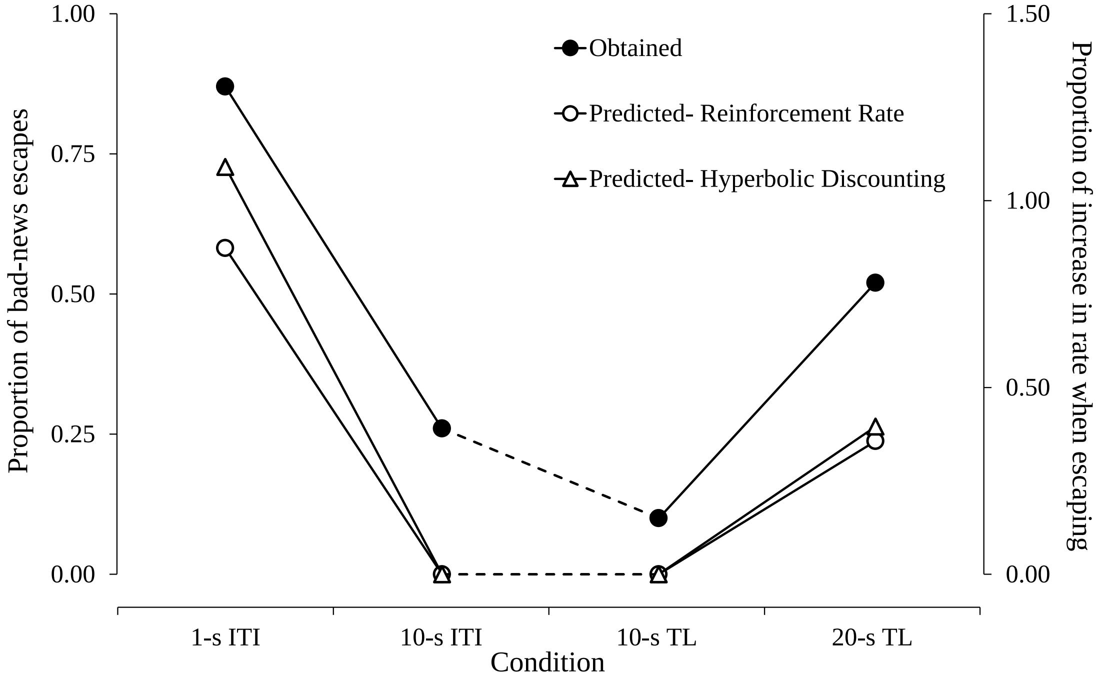


Figure 9. Mean proportion of observed escape responses (left y-axis) in each condition (black dots) and predicted increase in the reinforcement rate when escaping (right y-axis) according to the Reinforcement Rate Model (white dots) and the Hyperbolic Discounting Model (white triangles). The

filled symbols should be read on the left y-axis and the open symbols should be read on the right y-axis.

As Figure 9 shows, animals escaped more frequently from the bad-news stimulus when it was more profitable (Pearson's correlation between the proportion of escapes and the relative gain of escaping, $r = .98$, $p = .022$).

As previously noted, Experiments 1 and 2 shared a condition with a 10-s ITI and 10-s terminal links. Because all parameters were the same, the reinforcement rate when escaping and when not escaping was the same in both conditions, and, as a consequence, the gain of escaping was also the same (compare white dots in 10-s ITI and 10-s TL in Figure 9). Even though we initially expected pigeons to escape more frequently in the long box because moving away from the bad-news stimulus was congruent with moving towards the escape key, the opposite result was found: animals escaped more in the regular box (Experiment 1) than in the long box (Experiment 2, compare black dots in 10-s ITI and 10-s TL in Figure 9). One could hypothesize that this was because escaping in the long box was more effortful than in the regular box: not only birds had to peck the key, as in the regular box, but they also had to walk to the escape key. This explanation is, however, questionable because even when pigeons did not peck the escape key, they usually moved away from the bad-news stimulus towards the escape key.

Why then did pigeons almost always move towards the escape key, but only sometimes pecked at it? Even though we may look at the two responses, moving away from the bad-news stimulus and pecking the escape key, as being under control of the bad-news stimulus, this may not be the case. Possibly, animals moved away from the negative stimulus because in the natural environment the probability of encountering food increases if they move away from a patch with no food. Hence, moving away from the bad-news stimulus may be an unconditioned response to that stimulus, and for that reason it occurs in almost all trials. On the other hand, pecking the key may be controlled by the consequences of pecking, more specifically, by the change in the reinforcement rate caused by pecking. Possibly, this is why pecking only occurred when it was advantageous to do so.

In the present studies, to predict whether pigeons should escape the bad-news stimulus or not, we adapted the RRM of Vasconcelos et al. (2015) to incorporate the main characteristics of the prey model (Charnov, 1976). First, even though there is some evidence that animals ignore a negative stimulus when they are forced to be in its presence (e.g., Zentall, 2016), we assumed that when animals decide whether or not to reject a terminal-link stimulus, the duration of this stimulus is taken into account. Second, even though the RRM assumes that the ITI duration plays no role in the usual protocol (without an escape key), in the present task (with an escape key) the ITI was the immediate consequence of escaping, and therefore we assumed that escaping should be affected by the ITI duration. Third, to incorporate the prey model we assumed that animals would escape or accept the bad-news stimulus depending on which action yielded a higher reinforcement rate.

Even though our results were consistent with the RRM that assumes long-term rate maximization, they can also be explained by a short-term maximization process. The Hyperbolic Discounting Model (HDM; Mazur, 1984, 1987), a descriptive model that assumes short-term maximization of reinforcement rate, has been successful applied to a variety of situations, such as self-control choice (e.g., Bickel, Odum, & Madden, 1999; Fortes, Vasconcelos, & Machado, 2015), choice between fixed and variable schedules of reinforcement (e.g., Mazur, 1984, 2008), and choice between probabilistic reinforcers (e.g., Mazur, 1989; Myerson, Green, & Morris, 2011), both with human (e.g., Green, Fry, & Myerson, 1994) and non-humans animals (for a review, see Hayden, 2015).

The HDM describes how the strength or value of a reinforcer decreases as the delay to obtain it increases. According to this model, this decrease is hyperbolic, and it is given by

$$V = \sum_{i=1}^{n} P_i \left(\frac{A}{1 + KD_i} \right) \tag{7}$$

where V is the value of an alternative that delivers rewards after n different delays, with each delay D_i occurring with probability P_i. Importantly, D_i is not the total time between choice and reward delivery but the cumulative time in the presence of conditioned reinforcers. The total value of the alternative is a weighted average of the value of the possible rewards. The parameter A is related with the amount of reward, and the parameter K refers to how quickly the value of a reward decreases with delay. A $K = 1$ has been successfully used with pigeons (e.g., Mazur, 2005), so hereafter this value will be used.

Applying the HDM to the task used in these experiments and conjugating it with the prey model (thus allowing us to calculate $R_{\text{reject}} > R_{\text{accept}}$) leads to predictions very similar to those of the RRM (see Figure 9, white triangles). The correlation between the HDM's predictions and the observed data was also high and statistically significant, $r = .97$, $p = .028$. Even though the predictions of both models are very similar, they are not exactly the same. For example, when the ITI decreases from 10 to 1 s the HDM predicts a greater advantage in escaping than the RRM. Nonetheless, the results of the birds can be reasonably well predicted both by a model that assumes molar maximization (RRM) and a model that assumes molecular maximization (HDM).

The predictions derived from both models should be interpreted as only approximations to what the animal is expected to do, for the following reasons. First, we did not take into account the latencies to peck the keys, even though these latencies impose some delay to food. Second, in the HDM a value of $K = 1$ was used, which assumes that all intervals – terminal links signaled by different stimuli and the ITI - are discounted equally, while there is evidence that that might not be the case (e.g., Fortes et al., 2015). Third, both models predict an all-or-none pattern of escape responses from the bad-news stimulus, but we and others have consistently found intermediate values of escaping (the possible reasons for this are discussed elsewhere; e.g., Lea, 1979; Mazur, 2008; Shettleworth, 1988). For this reason, we predicted the proportion of escape responses in the bad-news stimulus using the relative increase in the value of escaping relative to not escaping, but other measures or decision rules can be applied.

To conclude, in the present experiments we changed a previously studied task to make it more similar to a natural foraging task. With this aim, we made information about the absence of food useful and animals responded appropriately using that information to increase food intake, which is consistent with the RRM's predictions. The major contribution of the present experiments was to show that animals do pay attention to the bad-news stimulus, but the presentation of that stimulus is probably not associated with the choice of the Informative Option. Because we were interested in modifying the original task the least, we introduced an escape key, thus making this complex task even more complex. In fact, to test whether pigeons should escape from a bad-news stimulus and, in the affirmative, under which circumstances, a simpler task should suffice: For example, instead of manipulating all ITIs, according to our analysis, only the duration of the ITI that follows rejection should matter. This and other assumptions (e.g., that animals take into account the bad-news duration when it is accepted) should be further explored. Moreover, we only tested two values of each variable of interest (the ITI and the terminal-link duration) and did not test values for which both models predict that *not* escaping is the optimal behavior. Future research should address these issues.

.

CHAPTER V

STUDY 3:

THE PARADOXICAL EFFECT OF LOWER REWARD PROBABILITY

Abstract

In a choice task, pigeons prefer an option that on 20% of the trials presents a stimulus always followed by food, and on the remaining 80% of the trials presents a stimulus never followed by food (Informative Option), over an option that provides food on 50% of the trials regardless of the stimulus presented (Non-informative Option). This preference has been attributed to the fact that, in the Informative Option, animals ignore (or do not engage with) the stimulus that is never followed by food. To assess whether (and when) pigeons would start attending to the stimulus that is usually never followed by food, we increased the probability of reinforcement in the presence of that stimulus in two experiments. Within the tested range, we found that the higher the probability of reinforcement, the less value the Informative Option has. To account for the results, we resorted to the Hyperbolic Discounting Model and the Reinforcement Rate Model, and derived the predicted results based on a linear and an all-or-none engagement function. The results were consistent with both models when applying a linear engagement function: The higher the probability of reinforcement following a stimulus, the higher the probability of engagement with that stimulus, and, paradoxically, the less value the option has.

Keywords: suboptimal choice, engagement, Hyperbolic Discounting Model, Reinforcement Rate Model, pigeons.

1. Introduction

No organism can attend to everything that happens around it. Instead, they selectively attend to the more relevant stimuli and ignore the ones unimportant for the current behavior. For example, when foraging, a squirrel may ignore the nearby rabbit, but it should not ignore the eagle flying above. In the laboratory, there is also evidence that animals do not pay attention to all stimuli or events occurring in an experimental session. For example, naïve birds may initially pay attention to the smell of the operant chamber or the sound of the experimenter coming into the room. However, as birds become more experienced, they start to ignore these events, paying more attention to the events relevant to the experimental task – the behavior of the animal is now under stimulus control.

There is, however, a task in which animals do seem to ignore an important event. In this task (Stagner & Zentall, 2010), pigeons choose between two options: the Informative Option that signals immediately whether food will be delivered or not, and the Non-informative Option that does not signal whether food is forthcoming or not (cf. initial-link, Figure 1). Specifically, the Informative Option delivers food after a 10-s delay on 20% of the trials, signaled by a red key (cf. terminal-link of Figure 1, stimulus S_1 with the subscript representing the probability of reinforcement following this stimulus); on the remaining 80% of the trials, signaled by a green key, pigeons wait 10 s but the trial always ends without food (S_p with $p = 0$, Figure 1). The Non-informative Option delivers food after 10 s on 50% of the trials, regardless of the signal presented (yellow – $S1_{0.5}$ – or blue – $S2_{0.5}$). Although the Non-informative Option provides food more often (50% vs. 20% of the trials), pigeons and starlings strongly prefer the Informative Option (Stagner, Laude, & Zentall, 2012; Stagner & Zentall, 2010; Vasconcelos, Monteiro, & Kacelnik, 2015; Zentall, 2016; Zentall & Stagner, 2011b). This result has been attributed to the fact that animals seem to ignore the S_p stimulus when $p = 0$, as explained below.

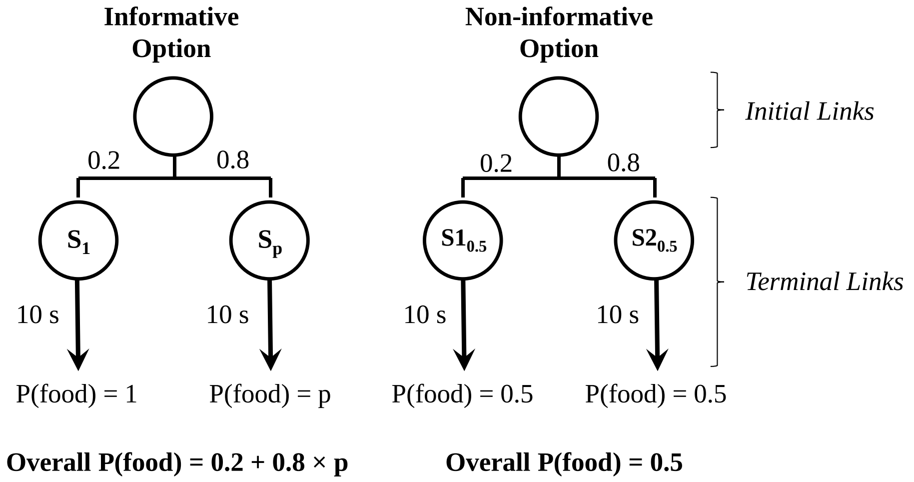


Figure 1. Design of Experiment 1. The probability of reinforcement following the S_p stimulus, p, was manipulated between subjects, and could be 0, 1/32, 6/32 or 12/32.

Two quantitative models have been put forth to explain performance in this task: the Hyperbolic Discounting Model, HDM (Mazur, 1984, 1987, 1997) and the Reinforcement Rate Model, RRM (Vasconcelos et al., 2015). Both models assume that this suboptimal behavior occurs because pigeons ignore the negative discriminative stimulus (S_p with $p = 0$, i.e., S_0), that is, its frequency and duration do not affect choice. This means that the Informative Option is perceived as always leading to S_1 followed by food, whereas the Non-informative Option is perceived as always leading to a stimulus that is intermittently followed by food. As a result, the probability of reinforcement in the Informative Option is perceived as higher than that of the Non-informative Option. Several experiments have confirmed this hypothesis by showing that changing the duration or probability of presentation of S_0 has nearly no effect on choice (Study 1; Mazur, 1995, 1996; Pisklak, McDevitt, Dunn, & Spetch, 2015; Spetch, Mondloch, Belke, & Dunn, 1994; Zentall, Laude, Stagner, & Smith, 2015). The clearest demonstration comes from an experiment in which we increased the duration of S_0 from 10 s up to 200 s, thus decreasing drastically the reinforcement rate in the Informative Option (Study 1) If animals were taking into account the time spent in the presence of S_0, the prediction was that they would start preferring the Non-informative Option as the S_0 duration increased. However, we found a 95% preference for the Informative Option at the longest delays, suggesting that the S_0 durations were not taken into account.

If, at steady state, animals do not pay attention to stimuli that are never followed by food[5] – but pay attention to stimuli that are always followed by food – one question that naturally arises is how the probability of attending to a stimulus and the probability of that stimulus being followed by food are related to each other; in other words, what is the engagement function. For instance, would it suffice to reinforce the S_p once per session, for the pigeon to take into account all the durations and episodes of non-reinforcement related to S_p? Note that, as an animal starts paying attention to a stimulus, it starts to consider not only the rewards (the perceived benefit of paying attention), but also all the time spent waiting for the reward (the perceived cost of paying attention). Without a model, it is not clear how to integrate the costs and benefits of paying attention to S_p, or how to integrate the value of the terminal links S_p and S_1 in order to compute the value of the Informative Option. In other words, computational models are needed to investigate how the value of the Informative Option changes as the probability p increases and the animal starts engaging with S_p.

As a first approach, we consider two possibilities for the engagement function: an all-or-none function, (i.e., a step function), and a continuous linear function. In both cases, the engagement function defines how the probability of taking into account a trial (i.e., its duration and reinforcer) varies as a function of the probability of reinforcement on that trial as signaled by the terminal-link stimulus. In the step function case, there is a threshold in the probability of reinforcement below which no trials are taken into account and above which all trials considered. In the linear function case, the engagement probability increases linearly as the probability of reinforcement increases. The main goal of the present study was to manipulate the probability of reinforcement p in the S_p stimulus, estimate the value of the Informative Option and test which engagement function better predicts the results. To that end, we modified the Hyperbolic Discounting Model (Mazur, 1984, 1987, 1997) and the Reinforcement Rate Model (Vasconcelos et al., 2015) to incorporate an engagement function. The predictions of each model with both engagement functions are discussed below.

---

[5] At the beginning of training, we have to assume that animals pay attention to all stimuli. In fact, there is evidence for that: in several experiments pigeons start the experiment showing a slight preference for the optimal, Non-informative Option (Laude, Stagner, & Zentall, 2014; Zentall & Stagner, 2011b). However, as they are exposed to the task, they may start to ignore and not engage with stimuli never followed by food (S_0). As a consequence, animals start to prefer the Informative Option.

1.1. Hyperbolic Discounting Model

According to the HDM (Mazur, 1984, 1987), the value of a delayed reward is given by

$$V = \frac{A}{1+KD} \tag{1}$$

where V is the value of the reward, A is related to the amount of food, D is the delay to the reward, and K is a discounting rate parameter related to how rapidly the value decreases as a function of delay. In most cases, setting K = 1 provides good estimates of value for pigeons (Mazur, 2005), so henceforth this value will be used. For example, according to Equation 1 (assuming A = 1, for simplicity), a reinforcer delayed by 10 s has V = 1/(1+10) ≈ 0.09 and a reinforcer delayed by 20 s has V = 1/(1+20) ≈ 0.005. Because the value decreases hyperbolically with delay, this model is called Hyperbolic Discounting Model.

Equation 1 applies when a reinforcer is delivered in every trial. However, in some tasks only a fraction of the trials are reinforced, and, as a consequence, the delay between choice and reward delivery is variable. For these cases, Mazur (1989) proposed that a probabilistic reinforcer is functionally equivalent to a reinforcer delivered after variable delays. Hence, in the case of probabilistic rewards, the value of a reward is given by

$$V = \sum_{i=1}^{n} P_i \left(\frac{A}{1+KD_i} \right) \tag{2}$$

where the value of a reinforcer, V, is the weighted average of the reinforcers delivered after n different delays, where P_i is the probability that a delay D_i will occur between choice and food delivery. Importantly, D_i does not include all durations between choice and food delivery. Instead, it only includes the time spent in the presence of conditioned reinforcers, that is, the duration of the stimulus presented before food, as clarified in the following example.

Consider, for instance, Option A that always presents a stimulus S for 10 s and ends with reinforcement on 20% of the trials. Because S is correlated with food to some extent, it is considered a conditioned reinforcer. The value of Option A depends on the cumulative duration of the conditioned reinforcer S. To illustrate, the reinforcer can be delivered on the first trial (with probability .2, after a delay of 10 s), on the second trial (with probability .16, after a cumulative delay of 20 s), or on the n^th^ trial with probability $.8^{n-1} \times .2$ after a cumulative delay of n × 10 s. The value of Option A is given by the weighted average of the value of all the discounted rewards: $V = .2/(1+10) + .16/(1+20) + \ldots + .8^{n-1} \times .2/(1+10n) \approx 0.04$. Consider now Option B: on 20% of the trials it is followed by stimulus S_1 that stays on for 10 s and ends with reinforcement, and, on the remaining trials, it is followed by a different stimulus, S_0, which is also presented for 10 s but is never reinforced. Because S_0 is never followed by food, it is not considered a conditioned reinforcer, and it does not enter the computations. In this case, Option B would have the same value of an option that always gives food, because that the parameters associated with the S_0 (e.g., probability of reinforcement and duration of presentation) are not taken into account: $V = 1/(1+10) \approx 0.09$. In a series of experiments, Mazur (1989, 1995, 1996; Mazur & Ratti, 1991) tested this model, and his predictions were generally confirmed.

Applying the HDM to the task depicted in Figure 1, (note the distinction between p, the probability of reinforcement in S_p and P_i, the probability that a delay D_i will occur) the Informative Option has the same value as an option that always delivers food after 10 s, $V_{Info} = 1/(1+10) \approx 0.09$, because all time spent in the presence of S_p is not taken into account when $p = 0$. On the contrary, in the Non-informative Option, both $S1_{0.5}$ and $S2_{0.5}$ are associated with reinforcement, so they are considered conditioned reinforcers and the duration spent in their presence is taken into account. For this option, the value is given by $V_{Non\text{-}Info} = .5\,[1/(1+10)] + .5^2 \times [1/(1+20)] + \ldots + .5^n \times [1/(1+ n \times 10)] \approx 0.06$. Because the value of the Informative Option is higher than the value of the other option (0.09 *vs.* 0.06), the model predicts a preference for the former.

In the example above, because $p = 0$, S_p was not attended to. However, if p were to increase, the animal would likely start paying attention to S_p. Next, we explore how the HDM can be adapted to accommodate two types of engagement functions: all-or-none and linear engagement functions.

1.1.1. All-or-None Engagement Function.

According to an all-or-none engagement function, if a stimulus is followed by reinforcement with probability equal or higher than a threshold θ, the stimulus becomes a conditioned reinforcer and in its presence the animal is always engaged in the task, taking into account both the durations and the reinforcers associated with that stimulus. If the stimulus is followed by reinforcement with probability lower than θ, the animal never engages. The engagement function $\varepsilon(p)$ can be defined as

$$\varepsilon(p) = \begin{cases} 0, & p < \theta \\ 1, & p \geq \theta \end{cases} \tag{3}$$

For the following analysis, let us assume that θ = .1, that is, any stimulus followed by reinforcement at least 10% of the times becomes a conditioned reinforcer and the animal engages with it. Then, in the Informative Option the animal always engages in S_1 trials (ε(1) = 1); whereas the animal may engage or not with in S_p, depending on the value of *p*: When *p* < .1, the animal never engages (ε(p) = 0), and when *p* ≥ .1, the animal always engages (ε(p) = 1). In the Non-informative Option, the animal is always engaged (i.e., ε(.5) = 1) because the probability of reinforcement (0.5) in both terminal links is higher than θ (0.1). Note that to calculate how the value of the Informative and Non-informative options changes with *p*, it is first necessary to determine, for each option, the probability that a given delay will occur.

Figure 2 shows a probability tree for all possible events. In the Informative Option, S_1 is presented on 20% of the trials, which always ends with food after a 10-s delay. This is an E_1 trial, with *E* meaning that the animal engages in these trials, and the subscript *1* meaning that the probability of reward in these trials is 1. On the remaining 80% of the trials, S_p is presented and the animal either engages and takes into account all of these trials (when ε(p) = 1), or does not engage (when ε(p) = 0; $\bar{E}$ trials). When it engages, with probability *p* it waits 10 s and receives food (E_p trials), and with probability *1 - p* it waits 10 s but no food is delivered (E_0 trials). The same reasoning applies to the Non-informative Option, but since the animal always engages in this option (ε(p) = 1), $\bar{E}$ trials never occur.

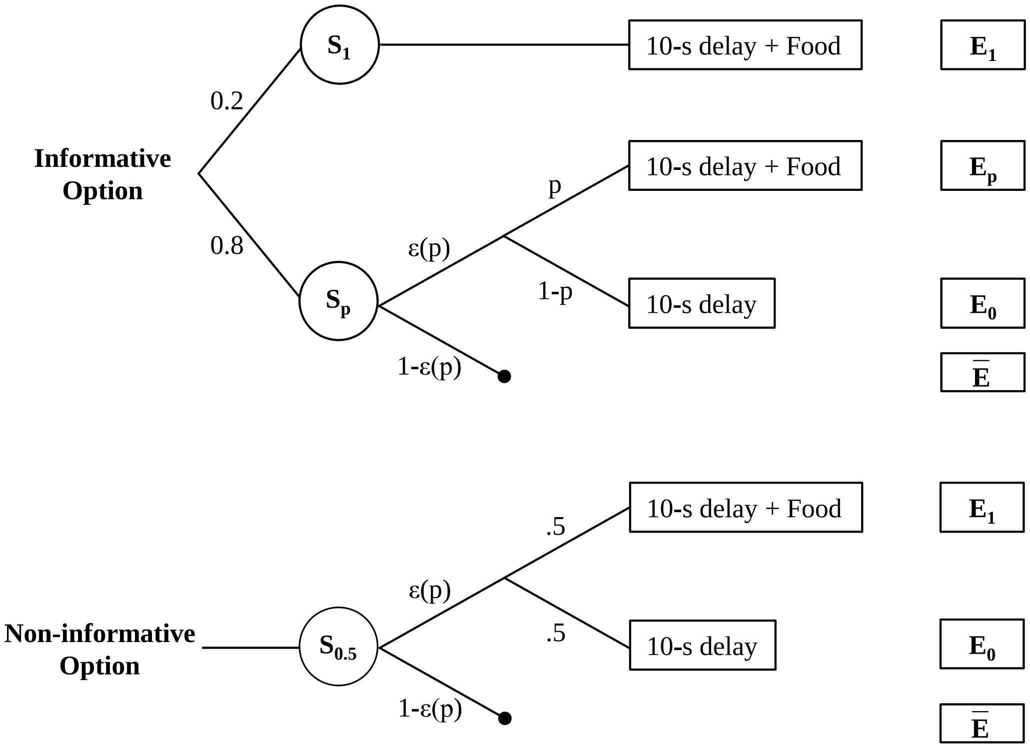


Figure 2. Probability tree for the possible events in the Informative and Non-informative options. Each possible type of trial was labeled according to whether the animal engaged or not (E_i or $\bar{E}$, respectively), and according to the probability of reinforcement in that trial, *i*.

Using the nomenclature of Figure 2, Appendix B presents the derivation of the value of each option. The computed values of each option are plotted in Figure 3. The top left panel shows the values of the Non-informative (dashed line) and Informative (dots) options as a function of *p* (the probability of reinforcement following S_p). For the Informative Option, both values for when the animal is engaged ($\varepsilon(p) = 1$, black dots) and disengaged ($\varepsilon(p) = 0$, white dots) are presented. For this analysis $K = 1$ and $D = 10$ s, but the relative position of the curves is preserved with different values of K and D.

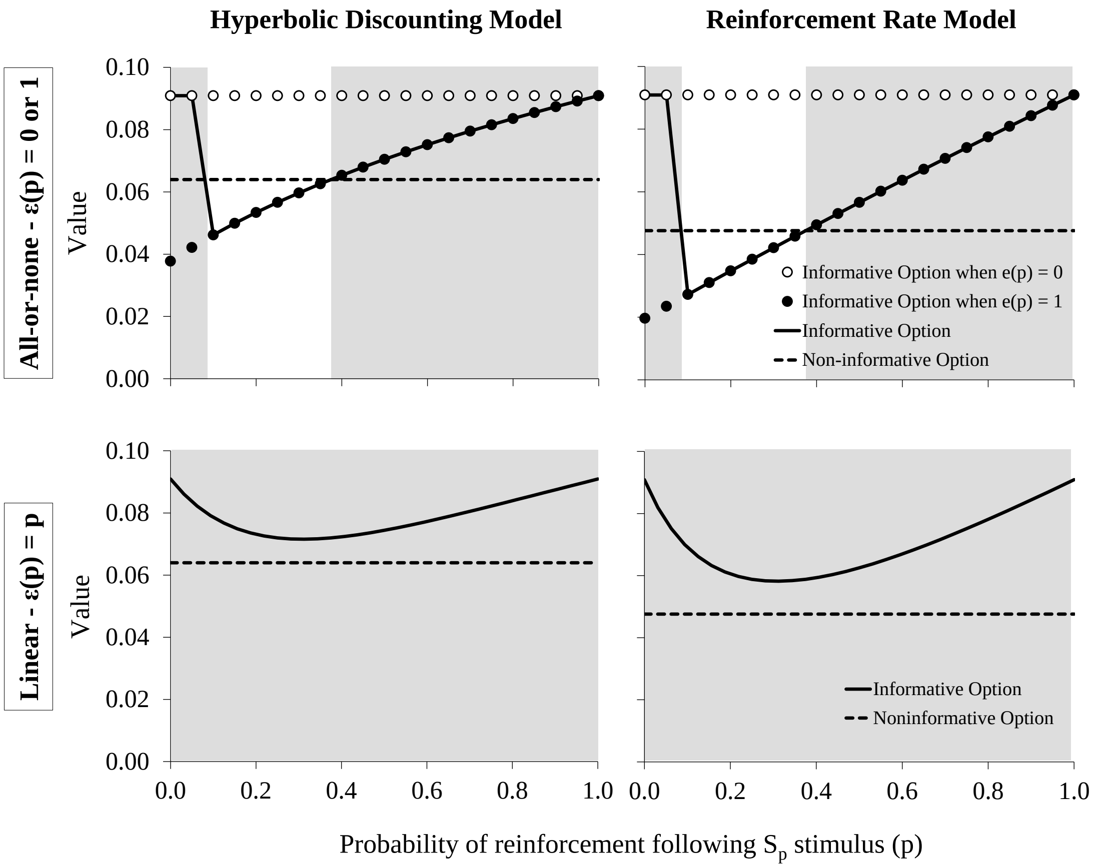


Figure 3. Predicted values of the Informative and Non-informative options as a function of the probability of reinforcement following S_p. The left panels show the predictions of the Hyperbolic Discounting Model with an all-or-none function (top left) and a linear engagement function (bottom left); the right panels show the predictions of the Reinforcement Rate Model also with an all-or-none function (top right) and a linear engagement function (bottom right). For the all-or-none predictions, a threshold $\theta = .1$ was used. A value of D = 10 s was used in all predictions, and for the Hyperbolic Discounting Model, K was set to 1. Choice of the Informative Option is predicted within the shaded areas.

Since the Non-informative Option is independent of *p*, its value remains constant. Conversely, the value of the Informative Option can vary greatly, depending on whether S_p is attended to or not. If the animal attends to S_p, as *p* increases, more reinforcers are delivered (and taken into account), so the value of the Informative Option (black dots) increases. If the animal never attends to S_p, the probability of food following that stimulus is never taken into account and the value of the Informative Option remains unchanged (white dots). The continuous line that connects the white and the black dots represents the actual value of the Informative Option, with a threshold $\theta = .1$.

When the S_p stimulus is never reinforced ($p = 0$), the probability of reinforcement is lower than the threshold ($p < \theta$), so the animal never engages with this stimulus. Therefore, at $p = 0$ the value of the Informative Option is given by the white dot. In this case (and while $p < .1$), the value of the Informative Option is higher than the value of the Non-informative Option (white dots vs. dashed line). However, when S_p is reinforced at least 10 % of the times ($p \geq .1$), the animal always engages in this task (i.e., $\varepsilon(p) = 1$) and the value of the Informative Option is given by the black dots. For instance, for $p = .1$, the value of the Informative Option is lower than the value of the Non-informative Option (black dot vs. dashed line). As p continues to increase, the value of S_p increases until it reaches a point where the Informative Option is perceived as more valuable than the Non-informative Option (the Informative and Non-informative options have the same value when $p = .375$).

In sum, if the engagement function assumes an all-or-none rule, animals should start preferring the Informative Option, but as p increases and surpasses θ, they should reverse preference, choosing more the Non-informative Option. Then, as p continues to increase and becomes higher than $p = .375$, animals should reverse preference again, choosing the Informative Option. Even though we used a threshold $\theta = .1$, this prediction holds for all $\theta < .375$, a reasonable threshold for engagement.

1.1.2. Linear Engagement Function.

The probability of reinforcement after a stimulus determines the strength of a stimulus as a conditioned reinforcer (Autor, 1969). In a linear engagement function, the strength of the conditioned reinforcer determines the likelihood of engagement: the stronger the conditioned reinforcer, the more likely is the animal to engage. The engagement function can be defined as

$$\varepsilon(p) = p \qquad (4)$$

which means that engagement is directly proportional to the probability of reinforcement: if a stimulus is always reinforced, the animal always engages with that stimulus, but if a stimulus is only reinforced half of the time the animal only engages in half of the trials, meaning that half of the reinforcers and durations are not be taken into account. This engagement function has implications in the probability that a given delay to food will occur (cf. Figure 2). The computation of the value of each option as a function of probability of reinforcement following S_p trials is shown in Appendix B.

The bottom left panel of Figure 3 shows the value of the Non-informative (dashed line) and the Informative (continuous line) options when $\varepsilon(p) = p$ as a function of p (the probability of reinforcement following S_P). For this analysis, K = 1 and D = 10 s, but different values of K and D preserve the relative position of the curves. As in the previous example, the value of the Non-informative Option is independent of S_P, so its value stays unchanged. Concerning the Informative Option, as the probability of reinforcement following S_P increases (and, as a consequence, as the overall probability of reinforcement in this option increases), its value follows a U-shaped function, with minimum at $p \approx .31$. For all values of *p,* the value of the Informative Option surpasses that of the Non-informative Option, thus predicting a preference for the former.

In sum, if the engagement function assumes a linear form, we expect the Informative Option to be preferred regardless of the probability of reinforcement in S_P. If the decision rule is a winner-takes-all rule, there should be an exclusive preference for the Informative Option, because it has a higher value than the Non-informative Option. However, even though the HDM (as the RRM) assume that the option with highest value should always be chosen, if two options have very similar values, they should to be difficult to discriminate. Therefore, although there could be a bias to choose the more valuable option, we should not always expect an exclusive preference. If we assume that what determines choice is the difference (or ratio) of the values of the two options (Baum & Rachlin, 1969), we expect the proportion of choices for the Informative Option to be always above .5, but have a shape similar to the value of the Informative Option: a high preference for the Informative Option in the lowest and highest values of *p,* but approaching indifference for intermediate values of *p*.

1.2. Reinforcement Rate Model

The RRM is based on optimal foraging principles and was developed by Vasconcelos and colleagues (2015) to account for the suboptimal preference found in the task depicted in Figure 1. According to this model, when given a choice, animals choose the option that maximizes the rate of food intake:

$$R = \frac{E}{T} \tag{5}$$

where R is the reinforcement rate of the option, E is the expected energy gain and T is total expected time spent to obtain that energy gain. If a given option yields reinforcement probabilistically, as in the present case, the reinforcement rate is given by

$$R = \frac{\sum_{i=1}^{n} P_i E_i}{\sum_{i=1}^{n} P_i T_i} \tag{6}$$

where the numerator is the expected probability of an energy gain (E) and the denominator is the expected delay to food (T). To be more precise, Equation 6 (with E = 1, for simplicity) can be decomposed into

$$R = \frac{P}{s + P \times (t + h) + (1 - P) \times t} \tag{7}$$

where a prey can be captured with probability P, and involves the following time costs. First, there is always a search time, s. Then, a prey is identified and the predator start pursuing it (h). With probability P, the prey is captured, and with probability $1\text{-}P$, the prey escapes. In the former case, the animal spends time pursuing the prey, t, and handling the prey once captured, h; in the latter case, only the pursuing time, t, is involved.

Applying Equation 7 to the present task, P is the probability of reinforcement in an option, s is the duration of the intertrial interval (ITI), t is the duration of the signals for food (S_1, S_p, $S2_{0.5}$ and $S2_{0.5}$; the equivalent of D in Equation 1), and h is the time of access to food. According to Equation 7, R increases monotonically with the probability of reinforcement associated to a given option, P. In its present form, however, Equation 7 ignores the informational imbalance of the task: Whereas the terminal links of the Informative Option convey information about whether or not reward is due soon (they are discriminative stimuli), the terminal links of the Non-informative option convey no information concerning reinforcement (they are nondiscriminative stimuli). In its final form, the RRM takes into account the usability of the information conveyed by the terminal links of the Informative Option: the information that a stimulus is always followed by food can be used differently than the information that a stimulus is never followed by food. To that end, two assumptions are necessary. First, when $p = 0$ as when the typical S_p is shown (i.e., when S_p signals that no food will be delivered), this stimulus is not taken into account. The reason for that comes from the natural foraging behavior of animals in the wild. While pursuing a prey, when an animal sees a signal indicating unambiguously that the prey will escape, the animal will most likely stop pursuing that prey and start searching for another one. For this reason, when the animal is presented with S_0 in the Informative Option, its durations are not included in the rate calculation. Second, because the same ITI follows all outcomes it and occurs after the reward has been delivered, it does not enter into the rate computation either (i.e., *s* is removed from Equation 7, too; for details see Vasconcelos et al., 2015, see also Bateson & Kacelnik, 1996; Mazur, 1989). Thus, assuming *h* = 1, the perceived rate on the Informative Option is given by

$$R_{Info} = \frac{P_{S_1}}{\cancel{s}+P_{S_1}\times\left(D+1\right)\cancel{+(1-P_{S_0})\times D}} = \frac{P_{S_1}}{P_{S_1}\times\left(D+1\right)} = \frac{1}{\left(D+1\right)} \qquad (8)$$

whereas the reinforcement rate on the Non-informative Option takes into account the two terminal-link stimuli, $S1_{0.5}$ and $S2_{0.5}$, because both are sometimes followed by food.

The predictions of the RRM when ε(p) follows an all-or-none function or a linear function, with D = 10 s, are shown next. The equations for the value of each option are presented in Appendix C.

1.2.1. All-or-None Engagement Function.

The top right panel of Figure 3 shows the value of the Non-informative (dashed line) and Informative (dots) options as a function of *p*, both when ε(p) = 0 (white dots) and ε(p) = 1 (black dots). The continuous line that connects the white and the black dots represents the actual value of the Informative Option, with a threshold θ = .1. Even though the values are not the same, the relative position of the Informative and Non-informative curves in the HDM and the RRM are similar: In both cases, as *p* increases, there should be an abrupt decrease in preference for the Informative Option followed by a steady increase.

1.2.2. Linear Engagement Function.

The bottom right panel of Figure 3 shows the value of the Non-informative (dashed line) and the Informative (continuous line) options when ε(p) = p, as a function of *p* (see Appendix C for the Equations). Again, the relative position of the Informative and Non-informative value functions are similar to the HDM: The Informative Option value decreases and then increases (with a minimum at $p \approx .31$), but it is always higher than the value of the Non-informative Option. Thus, according to a linear engagement function we should expect a preference for the Informative for all values of *p*. If the decision rule takes into account the relative value of the options, we should expect a stronger preference at the lowest and highest probabilities.

1.3. The Present Experiments

In the experiments that follow, we varied the probability of reinforcement in the presence of S_p from 0 to .375. At the higher limit, both the Informative and the Non-informative options yielded food on 50% of the trials. Note that, as *p* increases, the overall probability of reinforcement in the Informative Option also increases. Consequently, we could simply expect an increase in preference for the Informative Option as *p* increases. However, if animals engage with the stimulus according to the probability of reinforcement, as *p* increases, (at least up to a certain value) there should be a decrease in Informative Option value (see Figure 3; all predictions for the Informative Option show an initial decrease in value). The main difference between the all-or-none and linear engagement functions is that the first predicts a preference reversal as *p* increases, whereas the second does not. We conducted two experiments to test the all-or-none and the linear engagement functions. In Experiment 1, we measured *preference* for the Informative Option as a function of *p*. In Experiment 2, we used an adjusting-delay procedure to evaluate the *relative value* of the Informative Option as *p* increased.

2. Experiment 1 – Probability of food in the Informative Option

2.1. Method

2.1.1. Subjects

The subjects were eight pigeons (*Columba livia*) with previous experimental histories (mainly autoshaping and timing experiments). During the experiment the animals were maintained at about 80% of their free-feeding weights. In the colony room, grit and water was continuously available. The pigeons were housed in a room with controlled temperature (between 20º and 22º C) and light cycle (13h:11h light/dark, with lights on at 8 am).

2.1.2. Apparatus

Four Med Associates operant boxes for pigeons were used. The boxes were 28.5-cm high, 24-cm long and 30-cm wide. Each box was enclosed in a sound-attenuating chamber, equipped with a ventilation fan that also masked extraneous noises. The response panel had three circular keys, 2.5 cm in diameter. The keys were 6 cm apart (center-to-center) and the lowest edge was 21 cm above the floor grid. Also in the response panel, 4 cm above the floor grid there was a centrally aligned feeder opening (6-cm wide x 5-cm high). During reinforcement, the opening was illuminated with a 1.1-W light. In the panel opposite to the response panel a houselight (2.8 W) illuminated the whole box. This light was centrally located and 23 cm above the floor.

A personal computer with a custom Visual Basic 2010 program controlled the events and recorded data via Whisker software (Cardinal & Aitken, 2010).

2.1.3. Procedure

Pretraining. Two sessions of pretraining were conducted to make sure animals were pecking at all stimuli used in this task: red, green, yellow, blue and white on the left and right keys, and white on the center key. Each session had 44 trials, 4 trials with each stimulus. On each trial, one of these stimulus turned on, and one peck was reinforced with 3 s of access to food, followed by a 10-s ITI during which the houselight was on. In the second session, 10 pecks to each stimulus were required for food reinforcement.

Training. The sessions were composed of 120 trials, 40 forced-choice trials with each option (Informative and Non-informative) and 40 free-choice trials. The trials were randomly interspersed with the constraint that in each 30-trial block, 20 were forced-choice trials (10 per option) and 10 were free-choice trials. For half of the birds, the Informative and Non-informative options were always presented on the left and right keys, respectively; for the other half, the side allocation was reversed.

A trial began with the white center key flashing (250 ms on, 250 ms off). On forced-choice trials, one peck at the center key turned it off and one of the side keys was illuminated with a white hue (initial link, see Figure 1). One peck at the illuminated side key changed its color and a 10-s delay started. If the peck was at the Informative key, on 20% of the trials the key color changed to the S_1 stimulus (e.g., red), and after the 10-s delay the key turned off and food was delivered; on the remaining 80% of the trials the key changed to the S_p stimulus (e.g., green) and after the 10-s delay the key turned off and food was delivered with probability *p*. If the peck was at the Non-informative key, on 20% of the trials the key color changed to the $S1_{0.5}$ stimulus (e.g., yellow) and on the other 80% of the trials the key color changed to a different stimulus, $S2_{0.5}$ (e.g., blue). In both cases, the key stayed on for 10 s, after which food was delivered on 50% of the trials. The trials were separated by a 10-ITI with only the houselight on.

The terminal-link hues associated with each option were counterbalanced over pigeons, but red and green hues were always associated with one option and blue and yellow were always associated with the other option.

In free-choice trials the sequence of events was the same than in the forced-choice trials, except that after a center-key peck, both side keys were illuminated; One peck to one of the side keys turned off the other key and switched the selected key to one of the terminal hues as in the forced-choice trials.

Given that each session had 40 Informative forced-choice trials, 32 of those (i.e., 80%) were trials with the S_p stimulus. The probabilities of reinforcement following the S_p stimulus (p) were 0/32, 1/32, 6/32 or 12/32. Thus, when, for instance, $p = 1/32$ the bird was reinforced on the S_p stimulus in only one forced-choice trial, and possibly one time more in a free-choice trial. Note that when $p = 12/32$, for each 40 Informative trials, animals had 12 reinforcers following the S_p stimulus, and 8 reinforcers following the S_1 stimulus, so the overall probability of reinforcement was 50% in both the Informative and Non-informative options. The birds were divided into four groups each assigned to a p value. All birds received 15 sessions.

2.2. Results and Discussion

Figure 4 shows the percentage of choices for the Informative Option for each group across sessions. The acquisition data reveals that, for all probabilities of reinforcement in S_p, pigeons acquired a strong preference for the Informative Option: On the last three sessions all birds chose the Informative Option in all choice trials. The preference for the Informative Option regardless of the probability of reinforcement following the S_p stimulus does not support the all-or-none engagement function but is consistent with the predictions derived from the linear engagement function (cf. with bottom panels of Figure 3). Moreover, because an exclusive preference for the Informative Option was obtained, the decision rule does not seem to depend on the relative values of the options.

There were differences in the rate of preference acquisition depending on the probability of reinforcement on the Informative Option. Acquisition was fastest for group $p =$ 0/32, slowest for group $p = 12/32$, and intermediate for groups $p = 1/32$ and $p = 6/32$. A mixed analysis-of-variance (ANOVA) with session (15 levels) as the within-subjects factor and group (4 levels) as the between-subjects factor, confirmed a significant increase in preference for the Informative Option, as revealed by a significant effect of session, $F(14, 56) = 13.36$, $p < .001$, $\eta_G^2 = .64$. Even though there was not a significant effect of group, $F(3,4) = 0.63$, $p = .64$, $\eta_G^2 = .18$, the interaction between session and group was marginally significant, $F(42, 56) = 1.52$, $p = .071$, $\eta_G^2 = .37$, suggesting a possible difference between groups in acquisition of preference.

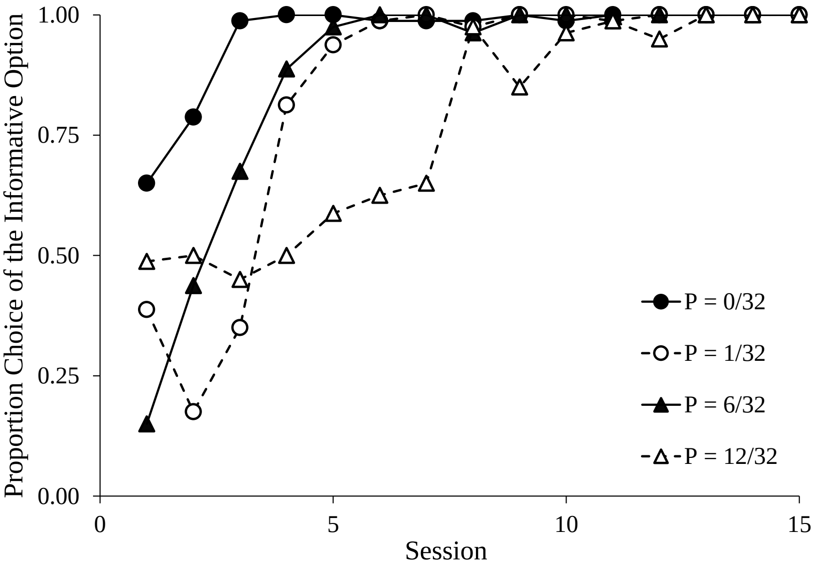


Figure 4. Mean proportion of choices of the Informative Option across sessions for each group in Experiment 1.

Although the results support the linear engagement function, they do not allow to discard the hypothesis that, as p increases, preference for the Informative Option increases simply because the overall probability of food in this option increases. As the probability of food following S_P increased, the overall probability of reinforcement in the Informative Option became more similar to the overall probability of reinforcement in the Non-informative Option: .20, .225, .35 or .50 versus .50 in the Non-informative Option. If the overall probability has some control over behavior, the slower acquisition for higher values of *p* may simply be explained by the fact that the discrimination between the perceived probabilities of food becomes more difficult (see Bailey & Mazur, 1990).

To conclude that the engagement function is linear, it does not suffice to show that the Informative Option is always preferred to the Non-informative Option. Instead, it is necessary to show that, even though the Informative Option had more value and was preferred, its value decreased with *p*. Although all birds showed a strong preference for the Informative Option, it does not mean that the Informative Option had the same value for all groups; it only means that for all groups the Informative Option had more value than the Non-informative Option. Thus, we may have obtained a ceiling effect which did not allow to see differences between groups. Experiment 2 addresses this issue.

3. Experiment 2 – Probability of food in the Informative Option with reduced delays in the Non-informative Option

Because all groups in Experiment 1 chose the Informative Option exclusively, it was not possible to detect differences in the Informative Option's value as *p* was manipulated. An alternative way to assess the value of the Informative Option is to increase the value of the Non-informative Option and see at which point subjects are indifferent between the two alternatives (Mazur 1987). Thus, in Experiment 2, we systematically decreased and then increased the terminal-link delays in the Non-informative Option (similarly to Zentall & Stagner, 2011b). As delays decrease, the Non-informative Option should become more attractive, and we expect a reversal in preference. On the contrary, as delays increase, the Non-informative Option should lose value and a preference for the Informative Option should arise again. The Non-informative terminal-link delays at which the animals are indifferent between the two options – the indifference point - can be interpreted as a measure of the value of the Informative Option. By comparing the indifference points from the same groups used in Experiment 1 we can assess the value of the Informative Option for each of them: If the Informative Option is highly valuable for one group, the Non-informative Option needs to become very attractive for the animal to start choosing it (i.e., its terminal-link delay needs to be quite short). In summary, the higher the value of the Informative Option, the lower the indifference point.

3.1. Method

3.1.1. Subjects

The subjects were the same as in Experiment 1.

3.1.2. Apparatus

The apparatus was the same as in Experiment 1.

3.1.3. Procedure

Experiment 2 started immediately after Experiment 1. The general procedure was the same as in Experiment 1, with the following differences. First, each session started with 80 randomly intermixed forced-choice trials (40 of each option), and then 80 free-choice trials followed, totaling 160 trials per session. Second, all terminal-link durations started at 20 s instead of 10 s. Finally, the terminal-link durations of the Non-informative Option were systematically decreased and then increased. The delays started at 20 s and were gradually decreased to 0 s in steps of 4 s. Then, the delays were increased from 0 s to 20 s, again in 4-s steps. Each duration was in effect for 4 sessions, totaling 44 sessions.

3.2. Results and Discussion

Figure 5 presents the proportion of choices of the Informative Option as a function of duration of the terminal links in the Non-informative Option. Each line represents a group (with a different probability of reinforcement following S_P), and each data point is the average of four sessions.

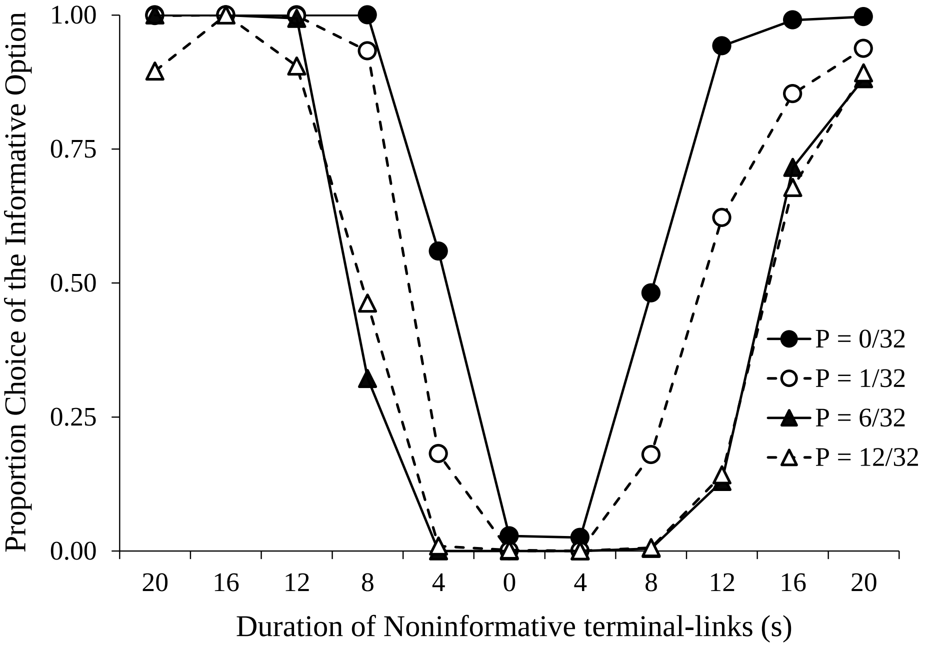


Figure 5. Mean proportion of choices of the Informative Option as a function of the non-informative terminal-link durations. Each data point averages two birds in four sessions.

Firstly, the leftmost data points show that all birds continued to prefer the Informative Option when all the terminal links were increased from 10 s (Experiment 1) to 20 s (Experiment 2). Secondly, j as the duration of the terminal links in the Non-informative Option decreased, all birds reversed their preference. Conversely, as the duration of the terminal links increased, the animals reversed preference again, now preferring the Informative Option. Moreover, the choice functions decreased and increased in an orderly way: As the Non-informative terminal-link delays decreased from 20 s to 0 s, preference for the Informative Option decreased; first for groups with a higher probability of reinforcement following S_P (p = 6/32 and p = 12/32; see triangles in Figure 5), then for group p = 1/32 (white dots) and finally for group p = 0/32 (black dots). Conversely, as the Non-informative terminal-link delays were increased from 0 s to 20 s, preference for the Informative Option increased; first for group p = 0/32, then for group p = 1/32 and finally for groups with a higher probability of reinforcement following S_P (p = 6/32 and p = 12/32). This symmetry in the choice functions shows that there is a positive correlation between the indifference points calculated from the ascending and the descending data, r = .87, p = .003. Although the curves are roughly symmetrical, they are slightly shifted to the right, suggesting some carry-over effects produced by the previous duration of the Non-informative terminal-links. This result is consistent with those from Zentall and Stagner (2011b).

To estimate the indifference point for each group, we averaged the descending and ascending indifference points (each calculated by linear interpolation from individual data). The results are presented in Figure 6; each white dot is an individual subject and the black, connected dots are the average for each group. The mean indifference point was 6.0 s for group p = 0/32, 8.7 s for group p = 1/32, 11.8 s for group p = 6/32 and 12.0 s for group p = 12/32. An one-way ANOVA revealed a significant effect of probability of reinforcement on the indifference points, $F(3, 4) = 16.92$, $p = .010$, $\eta_P^2 = .93$ with post-hoc Scheffe's tests revealing significant differences between group p = 0/32 and groups p = 6/32 and p = 12/32 (p = .019 and p = .017, respectively).

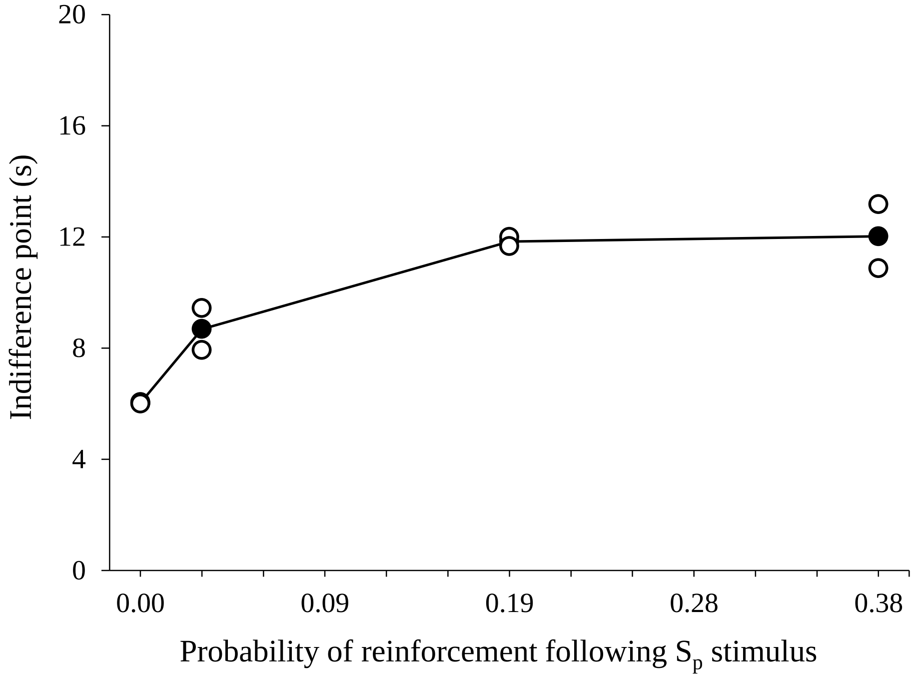


Figure 6. Indifference points (in seconds) as a function of the probability of reinforcement following the S_p stimulus. The white circles show the indifference points of individual birds and the black, connected circles show the mean indifference points

When p = 0/32 and a delay of 20 s, the Non-informative Option provided 2.5 times more reinforcement than the Informative Option and pigeons still preferred the latter; animals only reversed their preference when the terminal-link duration of the Non-informative Option was reduced by 70%. On the other extreme, for group p = 12/32, both options yielded food on half of the trials, but the non-informative terminal-link durations had to be reduced by 41% for the animals to reverse their preference. Put differently, even though the Informative

Option yielded more food for group p = 12/32 than for group p = 0/32, it required a smaller decrease in the non-informative terminal-link duration in group p = 12/32 for preference to reverse. This result suggests that, the higher the p, the less value the Informative Option has, which is consistent with a linear engagement function (cf. bottom panels of Figure 3).

Zentall and Stagner (2011b) also manipulated the non-informative terminal-link duration. In their experiment, the informative terminal links were 10-s long and p = 0 at all times. They found that animals were indifferent between the two alternatives when the Non-informative terminal-link durations were 4.4-s long, a reduction of 56%. In the present study, in a comparable situation (when p = 0/32), the delay had to be reduced by 70%, suggesting that in the present study the preference for the Informative Option was stronger than in Zentall and Stagner's (2011b). In fact, the indifference points are consistent with the preference data: whereas in the present study (Experiment 1 and first sessions of Experiment 2) all birds chose the Informative Option almost 100% of the trials, in Zentall's experiment choice stabilized around 80%. Although the reasons for this difference are not clear, it shows that the adjusting-delay procedure provides a reliable measure of the Informative Option value.

4. General Discussion

Animals attribute more value to an option that sometimes provides a signal perfectly correlated with the outcome (food or no food; the Informative Option) than to an option that provides signals uncorrelated with food (Non-informative Option), even when the latter provides more food (see, for instance, Stagner & Zentall, 2010; Mazur, 1995; Vasconcelos et al., 2015). The main goal of the present experiments was to assess how the value of the Informative Option changed as its probability of reinforcement increased. Based on the literature (Study 1; Mazur, 1995, 1996; Spetch et al., 1994; Stagner et al., 2012; Vasconcelos et al., 2015; Zentall et al., 2015), our premise was that this suboptimal preference occurs because, after learning the task, animals do not pay attention or do not engage with a stimulus that predicts the absence of reinforcement. By not engaging with such a stimulus, the Informative Option is akin to an option that always provides food and therefore its perceived value is higher than the other partially reinforced option. By increasing the probability of reinforcement following the until-then ignored stimulus (S_P in Figure 1), we expected that pigeons would start engaging with it (Beierholm & Dayan, 2010), and the question of interest was how this manipulation would affect the value of the Informative Option. We proposed two engagement functions, namely a linear engagement function, where the probability of engaging with the stimulus increases linearly with the probability of reinforcement, and an all-or-none threshold function, where animals only attend to the stimulus if the probability of reinforcement is at least equal to a given threshold. To predict the value of an option, we combined the aforementioned engagement function with two quantitative models able to predict preference in this task: the Hyperbolic Discounting Model (Mazur, 1984, 1987, 1997) and the Reinforcement Rate Model (Vasconcelos et al., 2015).

At a qualitative level the two models make the same predictions (see Figure 3). If the engagement is an all-or-none function (upper panels), the prediction is that animals should start preferring the Informative Option, but as its probability of reinforcement increases and reaches the threshold, the preference should reverse, that is, animals would suddenly start preferring the Non-informative Option. Increasing even more the probability of reinforcement should reverse preference again, at $p = .375$. In other words, in the range of probabilities tested in the present experiments - from 0 to 12/32 (or .375) - we predicted one reversal in preference as the probability of food increased. If, on the other hand, the engagement function is linear (lower panels) we predicted that, regardless of the probability of reinforcement, animals should prefer the Informative Option, even though the value of the two options should converge as the probability of food increased, up to a certain point. After that point, the value of the Informative option would increase and the difference in value between the two options should diverge again.

In Experiment 1, we measured preference for the Informative Option for different groups of birds exposed to different probabilities of reward in the Informative Option, and found that all groups preferred the Informative Option. This result is consistent with the linear engagement function. However, if the engagement function were indeed linear, the Informative Option value should be modulated by the probability of reinforcement. More specifically, if the decision rule is ratio- or difference-based, as the probability of reinforcement in the Informative Option increased, preference for that option should become less extreme within the tested range of probabilities.

In Experiment 2, we used an adjusting-delay procedure to test this prediction and found that the value of the Informative Option decreased as its probability of reward increased. The two experiments thus indicate that the more a stimulus is reinforced, the more the animals engage with it, indicating a linear engagement function. Even though other continuous monotonically increasing functions predict the same pattern of results, the main message is that the engagement function does not depend on a threshold, but is a continuous function of the probability of reinforcement. Although the results of our experiment do not allow us to unequivocally determine the nature of the engagement function, they do allow us to reject the all-or-none function.

Overall, our results are consistent with a linear engagement function, predicted both by the modified versions of Hyperbolic Discounting Model and the Reinforcement Rate Model. Despite their similarities, the models do not make the same predictions. Figure 7 shows, for each value of *p*, the average indifference point obtained (black dots), as well as the predicted indifference points from the HDM (with K = 1, white triangles) and the RRM (white dots), both with a linear engagement function. The figure shows that although both models predict the trend observed, they overestimate the indifference points. The fact that both models predict higher indifference points than those observed means that either the Informative Option has more value than predicted or the Non-informative Option has less value than predicted. Although the correlation between data and each model was the same (HDM: r = .97, *p* = .026; RRM: r = .97, *p* = .028), the RRM predicted indifference points closer to those observed. Thus, even though our main interest was to use the two models to predict the form of the engagement function and not to distinguish between the models, the RRM seems closer to our data.

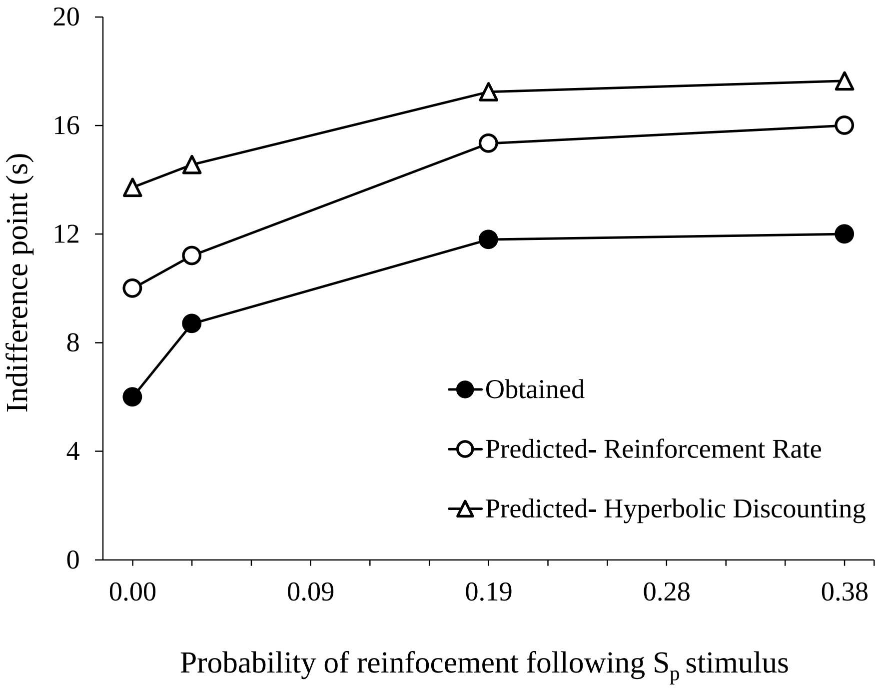


Figure 7. Obtained (black dots) and predicted indifference points by the Hyperbolic Discounting Model (white triangles) and the Reinforcement Rate Model (white dots) as a function of the probability of reinforcement following the S_P stimulus. A value of D = 20 s for the informative terminal-link delay was used in all predictions, and for the Hyperbolic Discounting Model K was set to 1.

What are the main similitudes and differences between the two models? Concerning the similitudes, both models assume that the value of an option depends on the rate of amount of food collected per unit of time. As a consequence, both models assume that the value of an option decreases hyperbolically as the delay to food increases. The major difference between the models rests in the rate computation. The HDM takes into account each possible delay to food (i.e., it considers the variability in the delay to reinforcement): Take for example ‘Option Var’ delivering food equiprobably after 1 or 19 s, and ‘Option Fix’ delivering food always after 10 s (the arithmetic mean of the delays in ‘Option Var’). For ‘Option Var’, the HDM computes the value of each delayed reinforcer, and then average the values. That is, it computes the average of the values of the two rewards, or the expectation of the ratios (EoR; for details see, Kacelnik & Bateson, 1996). Specifically (assuming K = 1), the value of Option Var would be V_{Var}=.5×[1/(1+1)] +.5×[1/(1+19)] = 0.275, more than the value of Option Fix, V_{Fix}=(1/(1+10) ≈ 0.090. The RRM, on the other hand, computes the average reward, the average delay to reward, and then computes their ratio, that is, it calculates the rate based on the ratio of the expectations (RoE; Kacelnik & Bateson, 1996). Contrarily to the HDM, for the RRM the value of an option is independent of the variability in the delay to food. Hence animals should be indifferent between the two alternatives whenever the average delay to food is the same. In this model, the rate of the Option Var would be $R_{Var}=1/(.5\times1+5\times19) = 0.1$, the same as the value of the Option Fix, $R_{Fix}=1/10 = 0.1$. There is considerable literature showing preference for variable delays over fixed delays with the same average, thus supporting the EoR models, such as the HDM (Mazur, 1984, 1986; Bateson & Kacelnik, 1996; however see Stephens & Krebs, 1986).

Because the rewards preceded by shortest delays convey greater value to an option, when we consider the variability in the delay to food (in the HDM), the inclusion of these more immediate rewards increase the value of the option. A consequence of this can be seen in the bottom panels of Figure 3, where we compare the value of each option for the HDM (left panel) and the RRM (right panel). As the probability of reinforcement following S_p increases, the impact on the value of the Informative Option is smaller in the HDM than in the RRM. This is due to the difference between the EoR and the RoE algorithms: The HDM takes into consideration all individually perceived rates, including rewards preceded by short delays (i.e., those yielding higher energy per time unit). Because the value of the Informative Option includes the rate perceived for individual rewards, its mean value is inflated, decreasing more slowly than in the RRM where only the average delay is considered.

A possible alternative explanation for the present results comes from Information Theory (Coombs, Dawes, & Tversky, 1970; Shannon, & Weaver, 1949). Broadly, Information Theory predicts that the higher the uncertainty of reward delivery associated with a stimulus, the less informative the stimulus is and the less value it has. Thus, a stimulus that signals a probability of reinforcement of 0 or 1 has maximum value (and minimum uncertainty) and a stimulus that signals a probability of reinforcement of 0.5 has minimum value (and maximum uncertainty, Shahan & Cunningham, 2015). In the present task, when the probability of reinforcement following S_p was zero, the Informative Option conveyed the maximum information possible because the outcomes of both stimuli were certain. As the probability of reinforcement following S_p started to increase, although the predictive value of S_1 remained at maximum value, the uncertainty associated with S_p started to increase. As a result, the Informative Option became less informative and its value decreased. This prediction is similar to the prediction of the HDM and the RRM with a linear engagement function, with the exception that these two models predict a minimum value at a probability around .31 (see Figure 3), whereas Information Theory predicts a minimum value at a probability around .50 (at the value of maximum uncertainty). Although the present study does not allow to discard the Information Theory hypothesis, this theory does not seem to explain the results of other experiments with this task.

For example, Roper and Zentall (1999) gave pigeons a choice between two options with the same probability of reinforcement. The Informative Option provided discriminative stimuli for food (i.e., one stimulus was always followed by food and the other was never followed by food) and the Non-informative Option provided nondiscriminative stimuli (i.e., uncorrelated with food). In different conditions the overall probability of reinforcement was kept equal between options, but was increased from 0.125 to 0.875. According to Information Theory, as the probability of food increases, the Informative Option maintains a high value because both stimuli convey the maximum amount of information. On the other hand, the uncertainty associated with the Non-informative Option increases up to p =.5 and then decreases (i.e., uncertainty follows an inverted U-shape). Thus, Information Theory predicts that, as the probability of reinforcement increases, preference for the Informative Option should also follow an inverted U-shape. Contrarily, Roper & Zentall (1999) found that preference slightly decreased as the probability increased. In another example, Zentall & Stagner (2011a) found that animals chose suboptimally even when there is no uncertainty associated with the options' outcomes. In their experiment, the Informative Option provided a stimulus followed by 10 pellets or a different stimulus followed by absence of food (0 pellets); the Non-informative Option provided a stimulus that always provided 3 pellets. Even though the information conveyed by all stimuli was maximal, pigeons preferred the Informative Option that sometimes provided a higher amount of food.

In conclusion, the results of the present experiments are yet another demonstration of suboptimal choice, with the additional counterintuitive finding that the more an option is reinforced, the less value it has. When the probability of food following S_P was zero, the Informative Option had *more value* than when the probability was greater than zero, suggesting that when $p = 0$ the delays to no food are not taken into account but some of those delays are included when the probability is increased. Moreover, we showed that the higher the probability of food, the higher the probability of engagement. We only tested probabilities equal to or lower than .375. If the engagement function is indeed linear, we should expect an increase in value for the Informative Option for higher probabilities of reinforcement. Future research should test a wider range of probabilities.

CHAPTER VI

CONCLUSION

In the present dissertation, we explored a task in which animals systematically behave suboptimally. We gave pigeons a choice between an Informative Option that provided discriminative stimuli (good- and bad-news) but was only reinforced 20% the time, and a Non-informative Option that did not provide discriminative stimuli but was reinforced 50% of the time. Even though the Non-informative Option provided 2.5 times more reinforcement, animals showed a strong, almost exclusive preference for the Informative Option. The major goal of this dissertation was to determine the role the bad-news stimulus plays in this task.

1. The effect of the bad-news stimulus on preference

Prior to the research reported here there was some evidence that manipulations of the frequency and duration of a bad-news stimulus have little to no effect on preference. To our knowledge, three studies have studied the effect of the frequency of the bad-news stimulus prior to us: Mazur (1996), Stagner, Laude, and Zentall (2012), and Vasconcelos, Monteiro, and Kacelnik (2015). For various reasons described in Study 1 (Chapter III), we thought that the evidence reported by these studies was not conclusive.

In Experiment 1 of Study I, we varied $p_{bad\ news}$ from 80% to 100% using a counterbalanced design. We found that when $p_{bad\ news} \leq 95\%$ animals strongly preferred the Informative Option, and when $p_{bad\ news} = 100\%$ animals strongly preferred the Non-informative Option. This result is more extreme than the obtained by Vasconcelos et al. (2015) with starlings: when $p_{bad\ news} = 0.95$, starlings were indifferent between the options, but in our study 5 out of 6 pigeons strongly preferred the Informative Option. RRM (Vasconcelos et al., 2015) predicts that animals should always prefer the Informative Option unless $p_{bad\ news} = 1$, which is consistent with our findings. However, when $p_{bad\ news} = 0.975$, we obtained mixed results: the early this condition was presented to the birds, the less they preferred the Informative Option. If one assumes that the most reliable results are those obtained when $p_{bad\ news} = 0.975$ was the first or second condition (because they were less affected by the previous conditions), then we conclude that animals prefer the Non-informative Option when $p_{bad\ news} = 0.975$. This is inconsistent with the predictions of the RRM and to account for it we need either to assume that a probability of 0.975 is virtually indistinguishable from 1.0 or to invoke a different process. For instance, choice of the Informative Option may require the probability of reinforcement on this option to be higher than a threshold, after which the probability of reinforcement does not affect preference. Still, this analysis is based on the results of only two birds, so caution is advised. The effect of higher probabilities of bad news should be further explored using a between-subject design.

Concerning the duration of bad-news stimulus, only Mazur (1995) and Spetch, Mondloch, Belke, and Dunn (1994) have studied its effect on preference. Mazur (1995) showed that increasing the bad-news duration from 30 s to 50 s did not affect the value of an option. Spetch et al. (1994) showed the same result when manipulating the bad-news duration from 5 s up to 55 s. In Experiment 2 of Study 1 (Chapter III), we further tested the boundary conditions for suboptimal preference by manipulating the bad-news duration. Since in Experiment 1 we showed that animals preferred the Non-informative Option when the probability of bad-news was high, we expected animals to also choose the Non-informative Option when the bad-news duration was very long. In this view, it was possible that previous studies did not find an effect of the bad-news duration because the maximum was 55 s, which may have not been long enough to decrease the value of the Informative Option. However, we found that even when the bad-news duration is 200 s, pigeons preferred the Informative Option around 95% of the time. From Study 1 we conclude that animals *ignore* the bad-news stimulus, in the sense that its probability and duration have little effect on the value of the Informative Option.

2. The test of escaping and the meaning of “ignoring”

Stating that animals *ignore* the bad-news stimulus as we described above, is nothing more than describing a result. In Study 2 (Chapter IV), we further explored the meaning of “ignoring the bad-news stimulus” by testing one of the main assumptions of the RRM. According to the model, the decision mechanism of animals is adapted to situations where they can use the information of the environment to change the current situation: for example, approach a good-news stimulus and escape from a bad-news stimulus. If this assumption is correct, when given the opportunity, pigeons should escape from bad news, but not from other stimuli in our task. That was indeed what we found: animals only escaped from bad news, and they escaped more when it was more advantageous to do so, a result consistent with other studies (e.g., Freidin, Aw, & Kacelnik, 2009). Moreover, even though animals escaped from the bad-news stimulus, they preferred almost exclusively the Informative Option.

Taken together, the escape and the preference results suggest that the bad-news stimulus exerted at least some control over behavior in Study 2. This conclusion is based on two arguments. First, the bad-news stimulus did not seem to influence the value of the Informative Option, because pigeons strongly preferred this option – even though the bad-news stimulus appeared 80% of the time. Second, animals escaped the bad-news stimulus and, when the terminal links were lengthened, they escaped even more. Thus, it seems that the bad-news stimulus did influence escape responding. We interpret these results in terms of the bad-news stimulus exerting control over escape behavior, but not being associated with the Informative Option. In other words, the bad-news stimulus does not influence preference for the Informative Option, but it influences escape behavior itself. From Study 2, we conclude that the RRM correctly predicts that animals escape from bad news and that animals ignore this stimulus in the sense that it is not attributed to the Informative Option.

3. Engaging with a stimulus and probability of reinforcement following the stimulus

Study 2 suggested the hypothesis that the presentation of a stimulus that is never followed by food - a bad-news stimulus - is not attributed to the choice of the Informative Option. In Study 3 (Chapter V), we asked what animals do when the bad-news stimulus starts being followed by food with a certain probability, p. On the one hand, increasing p increases the overall probability of reinforcement in the Informative Option, and therefore one could expect an increase in preference for this option (in case it is not prevented by a ceiling effect). On the other hand, if animals ignore the bad-news stimulus when $p = 0$ and start paying attention to it as p increases, then they should consider both the reinforcers and the time spent in the presence of this stimulus, which would lead to a devaluation of the Informative Option. We found that within the tested range (from $p = 0.0$ to 0.375), the higher the probability p, the less value the Informative Option had. This result is consistent with the hypothesis that the higher the probability of reinforcement following a stimulus, the more likely animals engage the stimulus. Based on the results of Study 2, we propose that the probability of reinforcement determines the association of the terminal-link stimulus with the initial-link option. From Study 3, we conclude that the probability of associating a stimulus with the Informative Option increases linearly with the probability of food following that stimulus.

4. The optimal foraging approach

In this dissertation we explored a task in which animals choose an option that yields less rather than more food. As animals choose suboptimally, performance in this task has been frequently signaled as an example eluding the theoretical and mathematical assumptions of optimal foraging theory (for recent reviews see Zentall, 2016; McDevitt, Dunn, Spetch, & Ludvig, 2016). The major challenge in the present work was to show that this suboptimal behavior can, indeed, be explained by a model based on optimal foraging principles.

In Study 2 and Study 3 we compared the predictions of the RRM (Vasconcelos et al., 2015) and the HDM (Mazur, 1984, 1987, 1997). In both cases the major trends of the data were equally well captured by the two models. The aim of comparing the models was not to choose one over the other, but rather to show that an optimal foraging model can account at least equally well for the data as a psychological model based on conditioned reinforcement. We believe that aim was accomplished.

Lastly, we emphasize that we do not see ecological and functional accounts in any sort of conflict with psychological or mechanistic accounts. On the contrary, we think that they complement each other. In this regard, we share Stevens' view:

> "A truly integrative study of decision making must synthesize evolutionary and psychological approaches. Though the emerging fields of cognitive ecology and evolutionary psychology have begun this integration, much work remains."
>
> (Stevens, 2011; p. 103)

REFERENCES

Abarca, N., & Fantino, E. (1982). Choice and Foraging. *Journal of the Experimental Analysis of Behavior, 38,* 117–123.

Abarca, N., Fantino, E., & Ito, M. (1985). Percentage reward in an operant analogue to foraging. *Animal Behaviour, 33,* 1096–1101.

Auge, R. J. (1974). Context, observing behavior, and conditioned reinforcement. *Journal of the Experimental Analysis of Behavior, 22*, 525–533.

Autor, S. M. (1969). The strength of conditioned reinforcers as a function of frequency and probability of reinforcement. In: Hendry, D.P. (Ed.), *Conditioned Reinforcement.* Homewood, IL: Dorsey Press.

Aw, J., Holbrook, R. I., Burt de Perera, T., & Kacelnik, A. (2009). State-dependent valuation learning in fish: Banded tetras prefer stimuli associated with greater past deprivation. *Behavioural Processes, 81,* 333-336.

Aw, J., Monteiro, T., Vasconcelos, M., & Kacelnik, A. (2012). Cognitive mechanisms of risky choice: Is there an evaluation cost? *Behavioural Processes, 89,* 95-103.

Bailey, J. T., & Mazur, J. E. (1990). Choice behavior in transition: Development of preference for the higher probability of reinforcement. *Journal of the Experimental Analysis of Behavior, 53,* 409-422.

Bateson, M., & Kacelnik, A. (1996). Rate currencies and the foraging starling: the fallacy of the averages revisited. *Behavioral Ecology, 7*, 341-352.

Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation. *Journal of the Experimental Analysis of Behavior, 12,* 861–874.

Bautista, L. M., Tinbergen, J., & Kacelnik, A. (2001). To walk or to fly? How birds choose among foraging modes. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 1089-1094.

Beierholm, U. R., & Dayan, P. (2010). Pavlovian-Instrumental interaction in “observing behavior.” *PLoS Computational Biology, 6*, e1000903.

Belke, T. W., & Spetch, M. L. (1994). Choice between reliable and unreliable reinforcement alternatives revisited: preference for unreliable reinforcement. *Journal of the Experimental Analysis of Behavior, 62,* 353–366.

Bickel, W. K., Odum, A. L., & Madden, G. J. (1999). Impulsivity and cigarette smoking: delay discounting in current, never, and ex-smokers. *Psychopharmacology, 146*, 447-454.

Blough, D. S. (2011). A random-walk model of accuracy and reaction time applied to three experiments on pigeon visual discrimination. Journal of Experimental Psychology: *Animal Behavior Processes, 37,* 133-150.

Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. *Trends in Cognitive Sciences, 11,* 118-125.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review, 113,* 700-765.

Bouton, M. E. (2007). *Learning and behavior: A contemporary synthesis*. Sunderland, MA: Sinauer Associates.

Bower, G., McLean, J., & Meacham, J. (1966). Value of knowing when reinforcement is due. *Journal of Comparative and Physiological Psychology, 62,* 184–192.

Browne, M. P., & Dinsmoor, J. A. (1974). Wyckoff's observing response: Pigeons learn to observe stimuli for free food but not stimuli for extinction. *Learning and Motivation, 5,* 165–173.

Brunner, D., Kacelnik, A., & Gibbon, J. (1992). Optimal foraging and timing processes in the starling, Sturnus vulgaris: effect of inter-capture interval. *Animal Behaviour, 44,* 597-613.

Cardinal, R. N., & Aitken, M. R. (2010). Whisker: a client-server high-performance multimedia research control system. *Behavior research methods, 42,* 1059-1071.

Charnov, E. L. (1976a). Optimal foraging: Attack strategy of a mantid. *The American naturalist, 110,* 141-151.

Charnov, E. L. (1976b). Optimal foraging: The marginal value theorem. *Theoretical Population Biology, 9,* 129-136.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical Psychology: An Elementary Introduction*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Cowie, R. J. (1977). Optimal foraging in great tits (Parus major). *Nature, 268,* 137-139.

Croze, H. (1970). Searching image in Carrion Crows. *Z. Tierpsychol., 5,* 1-85.

Cuthill, I. C., Haccou, P., & Kacelnik, A. (1994). Starlings (Sturnus vulgaris) exploiting patches: response to long-term changes in travel time. *Behavioral Ecology, 5,* 81-90.

Cuthill, I. C., Kacelnik, A., Krebs, J. R., Haccou, P., & Iwasa, Y. (1990). Starlings exploiting patches: the effect of recent experience on foraging decisions. *Animal Behaviour, 40,* 625-640.

Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An introduction to behavioural ecology* (4th ed.). Hoboken, NJ: Wiley-Blackwell.

Dinsmoor, J. A. (1983). Observing and conditioned reinforcement. *The Behavioral and Brain Sciences, 6,* 693–728.

Dinsmoor, J. A., Browne, M. P., & Lawrence, C. E. (1972). A test of the negative discriminative stimulus as a reinforcer of observing. *Journal of the Experimental Analysis of Behavior, 18*, 79–85.

Dinsmoor, J. A., Browne, M. P., Lawrence, C. E., and Wasserman, E. A. (1971). A new analysis of Wyckoff's observing response. *Proceedings of the 79th Annual Convention of the American Psychological Association*, 679-680.

Dinsmoor, J. A., Mueller, K. L., Martin, L. T., & Bowe, C. A. (1982). The acquisition of observing. *Journal of the Experimental Analysis of Behavior, 38*, 249–263.

Dunlap, A. S., & Stephens, D. W. (2009). Components of change in the evolution of learning and unlearned preference. *Proceedings of the Royal Society of London B: Biological Sciences, 276,* 3201-3208.

Dunn, R., & Spetch, M. L. (1990). Choice with uncertain outcomes: conditioned reinforcement effects. *Journal of the Experimental Analysis of Behavior, 53*, 201–218.

Fantino, E. (1969). Choice and rate of reinforcement. *Journal of the Experimental Analysis of Behavior, 12,* 723–730.

Fantino, E., & Abarca, N. (1985). Choice, optimal foraging, and the delay reduction hypothesis. *Behavioral and Brain Sciences, 8*, 315-330.

Fantino, E., & Preston, R. A. (1988). Choice and Foraging: The effects of accessibility and acceptability. *Journal of the Experimental Analysis of Behavior, 50*, 395–403.

Fantino, E., Dunn, R., & Meck, W. (1979). Percentage reinforcement and choice. *Journal of the Experimental Analysis of Behavior, 32,* 335–340.

Fortes, I., Vasconcelos, M., & Machado, A. (2015). The effect of response rate on reward value in a self-control task. *Journal of the Experimental Analysis of Behavior, 103*, 141–52.

Freidin, E., Aw, J., & Kacelnik, A. (2009). Sequential and simultaneous choices: Testing the diet selection and sequential choice models. *Behavioural Processes, 80,* 218-223.

Gibb, J. A. (1958). Predation by tits and squirrels on the eucosmid Ernarmonia conicolana (Heyl). *Journal of Animal Ecology, 27,* 375-396

Gibb, J. A. (1962a). L. Tinbergen's hypothesis of the role of specific search images. *Ibis, 104,* 106-111.

Gibb, J. A. (1962b). Tits and their food supply in English pine woods: a problem in applied ornithology. *Festschr. Vogelschutzwarte Hessen, Rheinland-Pfalz und Saarland,* 58-66.

Gibb, J. A. (1966). Tit predation and the abundance of Ernarmonia conicolana (Heyl.) on Weeting Heath, Norfolk, 1962-63. *Journal of Animal Ecology, 35,* 43-53.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review, 103,* 650.

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behavior*. New York: Oxford University Press.

Gipson, C. D., Alessandri, J., Miller, H. C., & Zentall, T. R. (2009). Preference for 50% reinforcement over 75% reinforcement by pigeons. *Learning & Behavior, 37,* 289–298.

Gloag, R., Fiorini, V. D., Reboreda, J. C., & Kacelnik, A. (2011). Brood parasite eggs enhance egg survivorship in a multiply parasitized host. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20112047.

Green, R. F. (1980). Bayesian birds: A simple example of Oaten's stochastic model of optimal foraging. *Theoretical Population Biology, 18*, 244-256.

Green, L., Fisher, E. B., Perlow, S., & Sherman, L. (1981). Preference reversal and self control: Choice as a function of reward amount and delay. *Behaviour Analysis Letters, 1,* 43-51.

Green, L., Fry, A. F., & Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison. *Psychological Science, 5*, 33-36.

Green, L., & Rachlin, H. (1977). Pigeons’ preferences for stimulus information: effects of amount of information. *Journal of the Experimental Analysis of Behavior, 27*, 255–263.

Hanson, J., & Green, L. (1989a). Foraging decisions: prey choice by pigeons. *Animal Behaviour, 37*, 429–443.

Hanson, J., & Green, L. (1989b). Foraging decisions: patch choice and exploitation by pigeons. *Animal Behaviour, 37,* 968–986.

Hayden, B. Y. (2015). Time discounting and time preference in animals: A critical review. *Psychonomic Bulletin & Review*, 1-15.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11-26.

Houston, A. I. (1986). The optimal flight velocity for a bird exploiting patches of food. *Journal of Theoretical Biology, 119,* 345-362.

Houston, A. I., & McNamara, J. M. (1999*). Models of adaptive behaviour: An approach based on state*. Cambridge: Cambridge University Press.

Houston, A., McNamara, J. M., & Steer, M. D. (2007). Do we expect natural selection to produce rational behaviour? *Philosophical Transactions of the Royal Society B, 362,* 1531–1543.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45*, 188–196.

Kacelnik, A. (1984). Central place foraging in starlings (Sturnus vulgaris). I. Patch residence time. *Journal of Animal Ecology, 53,* 283-299.

Kacelnik, A. (2003). The evolution of patience. In G. Loewenstein, D. Read, & R. Baumeister (Eds.), *Time and decision: Economic and psychological perspectives on intertemporal choice* (pp. 115-138). New York: Russell Sage Foundation.

Kacelnik, A. & Bateson, M. (1996). Risky theories: the effects of variance on foraging decisions. *American Zoologist*, *36*, 402–434.

Kacelnik, A., & Brunner, D. (2002). Timing and Foraging: Gibbon's Scalar Expectancy Theory and Optimal Patch Exploitation. *Learning and Motivation, 33,* 177-195.

Kacelnik, A., Cotton, P. A., Stirling, L., & Wright, J. (1995). Food allocation among nestling starlings: Sibling competition and the scope of parental choice. *Proceedings: Biological Sciences, 259*, 259-263.

Kacelnik, A., & Cuthill, I. (1987). Starlings and optimal foraging theory: modelling in a fractal world. In A. C. Kamil, J. R. Krebs, & H. R. Pulliam (Eds.), *Foraging Theory* (pp. 303-333). New York: Plenum Press.

Kacelnik, A., & Cuthill, I. (1990). Central place foraging in starlings (Sturnus vulgaris). II. Food allocation to chicks. *Journal of Animal Ecology, 59*, 655-674.

Kacelnik, A., & El Mouden, C. (2013). Triumphs and trials of the risk paradigm. *Animal Behaviour, 86*, 1117-1129.

Kacelnik, A., & Houston, A. I. (1984). Some effects of energy costs on foraging strategies. *Animal Behaviour, 32*, 609-614.

Kacelnik, A., & Todd, I. A. (1992). Psychological mechanisms and the Marginal Value Theorem: Effect of variability in travel time on patch exploitation. *Animal Behaviour, 43*, 313-322.

Kacelnik, A., Vasconcelos, M., Monteiro, T., & Aw, J. (2011). Darwin's "tug-of-war" vs. starlings' "horse-racing": How adaptations for sequential encounters drive simultaneous choice. *Behavioral Ecology and Sociobiology, 65*, 547-558.

Kagel, J. H., Battalio, R. C., & Green, L. (1995). *Economic choice theory: An experimental analysis of animal behavior*. Cambridge: Cambridge University Press.

Kagel, J. H., Green, L., & Caraco, T. (1986). When foragers discount the future: constraint or adaptation? *Animal Behaviour, 34*, 271-283.

Kendall, S. B. (1974). Preference for intermittent reinforcement. *Journal of the Experimental Analysis of Behavior, 21*, 463–473.

Kendall, S. B. (1985). A further study of choice and percentage of reinforcement. *Behavioural Processes, 10*, 399–413.

Krebs, J. R. (1973). Behavioral aspects of predation. In P. P. G. Bateson & P. Klopfer (Eds.), *Perspectives in Ethology* (pp. 73-111). New York: Plenum Press.

Krebs, J. R., Erichsen, J. T., Webber, M. I., & Charnov, E. L. (1977). Optimal prey selection in the great tit (Parus major). *Animal Behaviour, 25*, 30-38.

Laude, J. R., Beckmann, J. S., Daniels, C. W., & Zentall, T. R. (2014). Impulsivity affects suboptimal gambling-like choice by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 40*, 1–10.

Laude, J. R., Stagner, J. P., & Zentall, T. R. (2014). Suboptimal choice by pigeons may result from the diminishing effect of nonreinforcement. *Journal of Experimental Psychology: Animal Learning and Cognition, 40*, 12–21.

Lea, S. E. G. (1979). Foraging and reinforcement schedules in the pigeon: Optimal and non-optimal aspects of choice. *Animal Behaviour, 27*, 875–886.

Lieberman, D. A. (1972). Secondary reinforcement and information as determinants of observing behavior in monkeys (Macaca mulatta). *Learning and Motivation, 3*, 351-358.

Logue, A. W., Smith, M. E., & Rachlin, H. (1985). Sensitivity of pigeons to prereinforcer and postreinforcer delay. *Animal Learning and Behavior, 13*, 181–186.

Mazur, J. E. (1984). Tests of an equivalence rule for fixed and variable reinforcer delays. *Journal of Experimental Psychology: Animal Behavior Processes, 10*, 426-436.

Mazur, J. E. (1986). Fixed and variable ratios and delays: Further tests of an equivalence rule. *Journal of Experimental Psychology: Animal Behavior Processes, 12*, 116-124.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of* *behavio: The effect of delay and of intervening events on reinforcement value.* (Vol. 5, pp. 55–73). Hillsdale, NJ: Erlbaum.

Mazur, J. E. (1989). Theories of probabilistic reinforcement. *Journal of the Experimental Analysis of Behavior, 51*, 87–99.

Mazur, J. E. (1991). Choice with probabilistic reinforcement: effects of delay and conditioned reinforcers. *Journal of the Experimental Analysis of Behavior, 55*, 63–77.

Mazur, J. E. (1993). Predicting the Strength of a Conditioned Reinforcer: Effects of Delay and Uncertainty. *Current Directions in Psychological Science, 2*, 70–74.

Mazur, J. E. (1995). Conditioned reinforcement and choice with delayed and uncertain primary reinforcers. *Journal of the Experimental Analysis of Behavior, 63*, 139–150.

Mazur, J. E. (1996). Choice with certain and uncertain reinforcers in an adjusting-delay procedure. *Journal of the Experimental Analysis of Behavior, 66*, 63–73.

Mazur, J. E. (1997). Choice, delay, probability, and conditioned reinforcement. Animal *Learning & Behavior, 25*, 131–147.

Mazur, J. E. (2005). Effects of reinforcer probability, delay, and response requirements on the choices of rats and pigeons: possible species differences. *Journal of the Experimental Analysis of Behavior, 83*, 263–279.

Mazur, J. E. (2007). Choice in a successive-encounters procedure and hyperbolic decay of reinforcement. *Journal of the Experimental Analysis of Behavior, 88*, 73–85.

Mazur, J. E. (2008). Effects of reinforcer delay and variability in a successive-encounters procedure. *Learning & Behavior, 36*, 301–310.

Mazur, J. E. (2010). Distributed versus exclusive preference in discrete-trial choice. *Journal of Experimental Psychology: Animal Behavior Processes, 36*, 321-333.

Mazur, J. E., & Ratti, T. A. (1991). Choice behavior in transition: Development of preference in a free-operant procedure. *Animal Learning & Behavior, 19*, 241–248.

Mazur, J. E., & Romano, A. (1992). Choice with delayed and probabilistic reinforcers: Effects of variability, time between trials, and conditioned reinforcers. *Journal of the Experimental Analysis of Behavior, 58*, 513-525.

McDevitt, M. A., Dunn, R. M., Spetch, M. L., & Ludvig, E. A. (2016). When good news leads to bad choices. *Journal of the Experimental Analysis of Behavior, 105*, 23–40.

McDevitt, M. A., Spetch, M. L., & Dunn, R. (1997). Contiguity and conditioned reinforcement in probabilistic choice. J*ournal of the Experimental Analysis of Behavior, 68*, 317–327.

McDiarmid, C., & Rilling, M. (1965). Reinforcement delay and reinforcement rate as determinants of schedule preference. *Psychonomic Science, 2,* 195-196.

McNair, J. N. (1982). Optimal Giving-Up Times and the Marginal Value Theorem. *The American naturalist, 119,* 511-529.

McNamara, J. M., & Houston, A. I. (1985). Optimal foraging and learning. *Journal of Theoretical Biology, 117,* 231-249.

Mueller, K. L., & Dinsmoor, J. A. (1984). Testing the reinforcement properties of S-: A replication of Lieberman's procedure. *Journal of the Experimental Analysis of Behavior, 41*, 17–25.

Myerson, J., Green, L., & Morris, J. (2011). Modeling the effect of reward amount on probability discounting. *Journal of the Experimental Analysis of Behavior, 95,* 175-187.

Odum, A. L. (2011). Delay discounting: I'm a k, you're a k. *Journal of the Experimental Analysis of Behavior, 96,* 427-439.

Odum, A. L., & Baumann, A. L. (2003). Delay discounting: state and trait variable. In G. J. Madden & W. K. Bickel (Eds.), *Impulsivity: The Behavioral and Neurological Science of Discounting* (pp. 39-66). Washington, DC: American Psychological Association.

Parker, G. A. (1970). Sperm competition and its evolutionary effect on copula duration in the fly Scatophaga stercoraria. *Journal of Insect Physiology, 16*, 1301-1328.

Parker, G. A. (1978). Searching for mates. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (1st ed., pp. 214 -244). Oxford, UK: Blackwell Scientific Publications.

Parker, G. A., & Pizzari, T. (2010). Sperm competition and ejaculate economics. *Biological Reviews, 85,* 897-934.

Parker, G. A., Simmons, L. W., Stockley, P., McChristie, D. M., & Charnov, E. L. (1999). Optimal copula duration in yellow dung flies: effects of female size and egg content. *Animal Behaviour, 57*, 795-805.

Parker, G. A., & Stuart, R. A. (1976). Animal behavior as a strategy optimizer: Evolution of resource assessment strategies and optimal emigration thresholds. *The American naturalist, 110*, 1055-1076.

Pelé, M., & Sueur, C. (2013). Decision-making theories: linking the disparate research areas of individual and collective cognition. *Animal Cognition, 16,* 543-556.

Pisklak, J. M., McDevitt, M. A., Dunn, R. M., & Spetch, M. L. (2015). When good pigeons make bad decisions: Choice with probabilistic delays and outcomes. *Journal of the Experimental Analysis of Behavior, 104*, 241–251.

Pompilio, L., & Kacelnik, A. (2005). State-dependent learning and suboptimal choice: when starlings prefer long over short delays to food. *Animal Behaviour, 70*, 571-578.

Pompilio, L., Kacelnik, A., & Behmer, S. T. (2006). State-dependent learned valuation drives choice in an invertebrate. *Science, 311*, 1613-1615.

Prokasy, W. F. (1956). The acquisition of observing responses in the absence of differential external reinforcement. *Journal of Comparative and Physiological Psychology, 49*, 131–134.

Pulliam, H. R. (1974). On the theory of optimal diets. *The American naturalist, 108*, 59-74.

Rand, J. F. (1977). Behaviors observed during S- in a simple discrimination learning task. *Journal of the Experimental Analysis of Behavior, 27*, 103-117.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261-300.

Robinson, E. J. H., Franks, N. R., Ellis, S., Okuda, S., & Marshall, J. A. R. (2011). A Simple threshold rule is sufficient to explain sophisticated collective decision-making. *PLoS ONE, 6*, e19981.

Roper, K. L., & Zentall, T. R. (1999). Observing Behavior in Pigeons: The Effect of Reinforcement Probability and Response Cost Using a Symmetrical Choice Procedure. *Learning and Motivation, 30*, 201–220.

Samuelson, P. A. (1937). A note on measurement of utility. *The Review of Economic Studies, 4*, 155-161.

Seeley, T. D., Visscher, P. K., Schlegel, T., Hogan, P. M., Franks, N. R., & Marshall, J. A. R. (2012). Stop signals provide cross inhibition in collective decision-making by honeybee swarms. *Science, 335*, 108-111.

Shahan, T. A., & Cunningham, P. (2015). Conditioned reinforcement and information theory reconsidered. *Journal of the Experimental Analysis of Behavior, 103*, 405–418.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Shapiro, M. S., Siller, S., & Kacelnik, A. (2008). Simultaneous and sequential choice as a function of reward delay and magnitude: Normative, descriptive and process-based models tested in the European Starling (Sturnus vulgaris). *Journal of Experimental Psychology: Animal Behavior Processes, 34*, 75-93.

Shettleworth, S. (1988). The Psychology of Learning and Motivation: Advances in research and theory, 22. In: Bower, G. H. (Ed.), *Foraging as operant behavior and operant behavior as foraging: What have we learned* (pp. 1-50). San Diego: Academic Press.

Slagsvold, T., & Dale, S. (1991). Mate choice models: Can cost of searching and cost of courtship explain mating patterns of female pied flycatchers? *Ornis Scandinavica, 22*, 319-326.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27*, 161-168.

Smith, A. P., & Zentall, T. R. (2016). Suboptimal Choice in Pigeons : Choice Is Primarily Based on the Value of the Conditioned Reinforcer Rather Than Overall Reinforcement Rate. *Journal of Experimental Psychology: Animal Learning and Cognition, 42*, 1–9.

Sozou, P. D. (1998). On hyperbolic discounting and uncertain hazard rates. *Proceedings of the Royal Society of London B: Biological Sciences, 265*, 2015-2020.

Spetch, M. L., Belke, T. W., Barnet, R. C., Dunn, R., & Pierce, W. D. (1990). Suboptimal choice in a percentage-reinforcement procedure: effects of signal condition and terminal-link length. *Journal of the Experimental Analysis of Behavior, 53*, 219–234.

Spetch, M. L., Mondloch, M. V, Belke, T. W., & Dunn, R. (1994). Determinants of pigeons' choice between certain and probabilistic outcomes. *Animal Learning & Behavior, 22*, 239–251.

Stagner, J. P., Laude, J. R., & Zentall, T. R. (2011). Sub-optimal choice in pigeons does not depend on avoidance of the stimulus associated with the absence of reinforcement. *Learning and Motivation, 42*, 282–287.

Stagner, J. P., Laude, J. R., & Zentall, T. R. (2012). Pigeons prefer discriminative stimuli independently of the overall probability of reinforcement and of the number of presentations of the conditioned reinforcer. *Journal of Experimental Psychology: Animal Behavior Processes, 38*, 446–452.

Stagner, J. P., & Zentall, T. R. (2010). Suboptimal choice behavior by pigeons. *Psychonomic Bulletin & Review, 17*, 412–416.

Stephens, D. W., & Anderson, D. (2001). The adaptive value of preference for immediacy: when shortsighted rules have farsighted consequences. *Behavioral Ecology, 12*, 330–339.

Stephens, D. W., Brown, J. S., & Ydenberg, R. C. (2007). *Foraging: behavior and ecology*. Chicago: University of Chicago Press.

Stephens, D. W., & Dunlap, A. S. (2009). Why do animals make better choices in patch-leaving problems? *Behavioural Processes, 80*, 252–260.

Stephens, D. W., Kerr, B., & Fernández-Juricic, E. (2004). Impulsiveness without discounting: the ecological rationality hypothesis. *Proceedings of the Royal Society B, 271*, 2459–2465.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton, NJ: Princeton University Press.

Stephens, D. W., & McLinn, C. M. (2003). Choice and context: testing a simple short-term choice rule. *Animal Behaviour, 66*, 59–70.

Stevens, J. R. (2011). Mechanisms for decisions about the future. In R. Menzel, & J. Fischer (Eds.). *Animal thinking: Contemporary issues in comparative cognition* (pp. 93-104). Cambridge, MA: MIT Press.

Stevens, J. R., & Stephens, D. W. (2010). The adaptive nature of impulsivity. In G. J. Madden & W. K. Bickel (Eds.), Im*pulsivity: The Behavioral and Neurological Science of Discounting* (pp. 361-387). Washington, DC: American Psychological Association.

Stevens, J. R., Rosati, A. G., Ross, K. R., & Hauser, M. D. (2005). Will travel for food: Spatial discounting and reward magnitude in two New World monkeys. *Current Biology, 15*, 1855-1860.

Templeton, A. R., & Lawlor, L. R. (1981). The fallacy of the averages in ecological optimization theory. *The American naturalist, 117,* 390-393.

Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift für Tierpsychologie*, *20,* 410-433.

Turelli, M., Gillespie, J. H., & Schoener, T. W. (1982). The fallacy of the fallacy of the averages in ecological optimization theory. *The American naturalist, 119*, 879-884.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* S251–S278.

Vasconcelos, M., Carvalho, M. P., & Machado, A. (in press). Timing in animals from the natural environment to the laboratory, from data to models. In J. Call (Ed.), *APA Handbook of Comparative Psychology*. Washington, DC: American Psychological Association.

Vasconcelos, M., & Urcuioli, P. J. (2008). Deprivation level and choice in pigeons: A test of within-trial contrast. *Learning and Behavior, 36*, 12-18.

Vasconcelos, M., Monteiro, T., Aw, J., & Kacelnik, A. (2010). Choice in multi-alternative environments: A trial-by-trial implementation of the Sequential Choice Model. *Behavioural Processes, 84*, 435-439.

Vasconcelos, M., Monteiro, T., & Kacelnik, A. (2013). Context-dependent preferences in starlings: Linking ecology, foraging and choice. *PLoS ONE, 8*, e64934.

Vasconcelos, M., Monteiro, T., & Kacelnik, A. (2015). Irrational choice and the value of information. *Scientific Reports, 5*, 13874.

Ward, R. D., Gallistel, C. R., & Balsam, P. D. (2013). It's the information! *Behavioural Processes, 95*, 3–7.

Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.

Wright, J., & Cuthill, I. (1990). Biparental care: short-term manipulation of partner contribution and brood size in the starling, Sturnus vulgaris. *Behavioral Ecology, 1*, 116-124.

Wyckoff, L. B. (1969). The role of observing responses in discrimination learning. In: Hendry, D. (Ed.), *Conditioned reinforcement* (pp. 237-260). Homewood, IL: The Dorsey Press.

Zentall, T. R. (2014). Suboptimal choice by pigeons: An analog of human gambling behavior. *Behavioural Processes, 103*, 156–164.

Zentall, T. R. (2016). Resolving the Paradox of Suboptimal Choice. *Journal of Experimental Psychology: Animal Learning and Cognition, 42*, 1-14.

Zentall, T. R., Laude, J. R., Stagner, J. P., & Smith, A. P. (2015). Suboptimal Choice by Pigeons: Evidence that the Value of the Conditioned Reinforcer Rather than its Frequency Determines Choice. *The Psychological Record*, *65*, 223–229.

Zentall, T. R., & Stagner, J. P. (2011a). Maladaptive choice behaviour by pigeons: an animal analogue and possible mechanism for gambling (sub-optimal human decision-making behaviour). *Proceedings of the Royal Society B, 278*, 1203–1208.

Zentall, T. R., & Stagner, J. P. (2011b). Sub-optimal choice by pigeons: Failure to support the Allais paradox. *Learning and Motivation, 42*, 245–254.

APPENDICES

APPENDIX A

Individual preferences for the Informative Option in the last three sessions of the Adjusting-delay Condition and in the Fixed-delay Condition.

Condition	30-trial block	Pigeon								Mean
		P463	PG13	P665	P022	P876	P748	P948	PG35	
Adjusting-delay	-2	1.00	0.90	0.90	1.00	1.00	1.00	0.90	0.90	0.95
	-1	1.00	1.00	1.00	1.00	0.70	1.00	0.90	0.80	0.93
	0	1.00	1.00	1.00	1.00	0.90	1.00	0.90	0.80	0.95
Fixed-delay	1	1.00	1.00	1.00	1.00	0.60	1.00	0.80	0.20	0.83
	2	1.00	1.00	1.00	1.00	0.70	1.00	0.90	0.50	0.89
	3	1.00	1.00	1.00	1.00	0.80	0.80	1.00	0.70	0.91
	4	1.00	1.00	1.00	1.00	0.80	0.90	0.90	0.40	0.88
	5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.99
	6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.60	0.95
	7	1.00	1.00	1.00	1.00	0.90	1.00	1.00	0.70	0.95
	8	1.00	1.00	0.90	0.90	1.00	1.00	1.00	0.90	0.96
	9	1.00	1.00	0.70	1.00	1.00	0.90	1.00	1.00	0.95

APPENDIX B

1. Hyperbolic Discounting Model

To calculate the value of an option we first need to compute the probability of a delay D_i to reward. Then, we apply Equation 2, which gives a weighted average of the value of all delayed rewards. A value of $A = 1$ is used.

1.1. All-or-None Engagement Function

For the Informative Option we consider two possibilities, ε(p) = 0 (S_p is never attended to) and ε(p) = 1 (S_p is always attended to). Note that, in the Informative Option, S_1 is always attended to (cf. Figure 2). For the Non-informative Option we consider ε(p) = 1 (both $S1_{0.5}$ and $S2_{0.5}$ are always attended to).

1.1.1. Value of the Informative Option

In the Informative Option, when ε(p) = 0, all S_p trials are $\overline{\mathrm{E}}$ trials, that is, neither the time nor the reinforcers associated with this stimulus are taken into account (cf. Figure 2). As a consequence, there are only two types of trials: E_1 and $\overline{\mathrm{E}}$. However, only the E_1 trials are attended to, so the perceived delay to food is always D s. Thus, when ε(p) = 0 the value of the Informative Option (with $A = 1$) is given by:

$$V_{Info\ \varepsilon=0} = \frac{1}{1 + K \times D} \tag{A1}$$

On the other hand, when ε(p) = 1, every trial is attended to, so $\overline{\mathrm{E}}$ trials never occur. Thus, there are three types of trials: E_1, E_p and E_0. The outcome of each trial can be classified either as a success (when food is delivered, in E_1 and E_p trials) or a failure (when food is not delivered, in E_0 trials). To calculate the probability of each delay to food we can simply calculate the probability of having *i* trials to food, that is, *i - 1* failures followed by a success. Because each trial adds D s, the total delay to food equals (number of trials to food) × (D). Let *X* equal the number of trials until a reward is delivered (including the trial with reward). The probability mass function of *X* follows a geometric distribution

$$P\{X=i\}=\left(P(\mathit{failure})\right)^{i-1}P(\mathit{success}) \qquad i=1,\ 2,\ \ldots \tag{A2}$$

When $\varepsilon(p) = 1$ the successes occur with probability $P(E_1) + P(E_p)$, and the failures occur with probability $P(E_0)$. Substituting in Equation A2 we obtain

$$P\{X=i\}=\left(P(E_0)\right)^{i-1}\left(P(E_1)+P(E_p)\right) \qquad i=1,\ 2,\ \ldots \tag{A3}$$

Thus, the probability of having a delay $i \times D$ to food is given by

$$P(iD)=\left(P(E_0)\right)^{i-1}\left(P(E_1)+P(E_p)\right) \qquad i=1,\ 2,\ \ldots \tag{A4}$$

The value of the Informative Option when $\varepsilon(p) = 1$ is then computed according to Equation 2 and Equation A4:

$$\begin{aligned} V_{Info\ \varepsilon=1} &= \sum_{i=1}^{n}\frac{P(E_0)^{i-1}\left(P(E_1)+P(E_p)\right)}{1+K\times iD} \\ &= \left(P(E_1)+P(E_p)\right)\sum_{i=1}^{n}\frac{P(E_0)^{i-1}}{1+K\times iD} \end{aligned} \tag{A5}$$

With the help of the probability tree in Figure 2, we calculated the values of $P(E_1)$, $P(E_p)$ and $P(E_0)$, and substituted in the Equation A5, obtaining

$$V_{Info\ \varepsilon=1}=(.2+.8p)\sum_{i=1}^{n}\frac{\left(.8\times(1-p)\right)^{i-1}}{1+K\times iD} \tag{A6}$$

1.1.2. Value of the Non-Informative Option

The same reasoning applies to calculate the probability of each delay in the Non-informative Option with $\varepsilon(p) = 1$. The probability of success is $P(E_1)$ and failure is $P(E_0)$. The value of this option is given by

$$V_{Non-info\ \varepsilon=1}=\sum_{i=1}^{n}\frac{P(E_0)^{i-1}\times P(E_1)}{1+K\times iD} \tag{A7}$$

That simplifies to

$$\begin{aligned} V_{Non-info\ \varepsilon=1} &= \sum_{i=1}^{n}\frac{.5^{i-1}\times .5}{1+K\times iD} \\ &= \sum_{i=1}^{n}\frac{.5^{i}}{1+K\times iD} \end{aligned} \tag{A8}$$

when the used probabilities are inserted.

1.2. Linear Engagement Function

The value of each option is calculated assuming ε(p) = *p*.

1.2.1. Value of the Informative Option

In the Informative Option, when ε(p) = *p*, there are four types of trials: E_1, E_p, E_0 and $\overline{E}$. These trials can be divided into successes (E_1 and E_p trials) and failures (E_0 and $\overline{E}$ trials). Because we are only interested in the trials that add a delay to food, the $\overline{E}$ trials are irrelevant and can be ignored. Our sample space can then be reduced to $S = \{E_1, E_p, E_0\}$. Thus, the probability of success and failure are given by

$$P_{Info\ \varepsilon=p}(success) = \frac{P(E_1)+P(E_p)}{P(E_1)+P(E_p)+P(E_0)} \qquad \text{(A9)}$$

$$P_{Info\ \varepsilon=p}(failure) = \frac{P(E_0)}{P(E_1)+P(E_p)+P(E_0)} \qquad \text{(A10)}$$

It follows that the probability of a delay *i* × *D* to food is given by

$$P(iD) = \left(\frac{P(E_0)}{P(E_1)+P(E_p)+P(E_0)}\right)^{i-1}\left(\frac{P(E_1)+P(E_p)}{P(E_1)+P(E_p)+P(E_0)}\right) \qquad i=1,\ 2,\ \ldots \qquad \text{(A11)}$$

The value of the Informative Option when ε(p) = *p* is then

$$V_{Info\ \varepsilon=p} = \frac{P(E_1)+P(E_p)}{P(E_1)+P(E_p)+P(E_0)}\sum_{i=1}^{n}\frac{\left(\frac{P(E_0)}{P(E_1)+P(E_p)+P(E_0)}\right)^{i-1}}{1+K\times i\times D} \qquad \text{(A12)}$$

Applying the used probabilities and assuming ε(p) = *p* we obtain

$$\begin{aligned} V_{Info\ \varepsilon=p} &= \frac{.2+.8p^2}{.2+.8p^2+.8p(1-p)}\sum_{i=1}^{n}\frac{\left(\frac{.8p(1-p)}{.2+.8p^2+.8p(1-p)}\right)^{i-1}}{1+K\times i\times D} \\ &= \frac{.2+.8p^2}{.2+.8p}\sum_{i=1}^{n}\frac{\left(\frac{.8p-.8p^2}{.2+.8p}\right)^{i-1}}{1+K\times i\times D} \end{aligned} \qquad \text{(A13)}$$

1.2.2. Value of the Non-informative Option

A similar reasoning applies to compute the probability that a given delay to food will occur in the Non-informative Option. The main difference between calculating the value of the Informative Option and the Non-informative Option when ε(p) = *p* is that, in the second case there is only one type of success: the E_1 trials.

The probability of a delay $i \times D$ to food, with $i - 1$ failures followed by a success is given by

$$P(iD) = \left(\frac{P(E_0)}{P(E_1) + P(E_0)} \right)^{i-1} \left(\frac{P(E_1)}{P(E_1) + P(E_0)} \right) \qquad i = 1,\ 2,\ \ldots \tag{A14}$$

The value of the Non-informative Option when ε(p) = *p* is then

$$V_{Non-info\ \varepsilon=p} = \frac{P(E_1)}{P(E_1) + P(E_0)} \sum_{i=1}^{n} \frac{\left(\frac{P(E_0)}{P(E_1) + P(E_0)} \right)^{i-1}}{1 + K \times i \times D}$$

(A15)

In this option, $p = .5$, and therefore ε(p) = .5 (see Figure 2). Using these values to compute the probabilities of each type of trial, we obtain

$$\begin{aligned} V_{Non-info\ \varepsilon=p} &= \frac{.25}{.25 + .25} \sum_{i=1}^{n} \frac{\left(\frac{.25}{.25 + .25} \right)^{i-1}}{1 + K \times i \times D} \\ &= .5 \sum_{i=1}^{n} \frac{(.5)^{i-1}}{1 + K \times i \times D} \end{aligned} \tag{A16}$$

Interestingly, Equation A16 is equivalent to the value of the Non-informative Option when the engagement function is all-or-none (that is, ε(p) = 1; see Equation A8):

$$V_{Non-info\ \varepsilon=p} = V_{Non-info\ \varepsilon=1} = \sum_{i=1}^{n} \frac{.5^{i}}{1 + K \times iD} \tag{A17}$$

The equivalence between equations A8 and A16 (when ε(.5) = 1 and when ε(.5) = .5, respectively) can be easily understood if one thinks that, from the animal's viewpoint, the relative proportion of reinforcers and delays is the same when it engages in all trials or engages in only half of the trials: in both cases the animal receives food on half of the engaged trials. In other words, the value of the Non-informative Option is influenced by the probability of reinforcement but not by the probability of engagement.

APPENDIX C

1. Reinforcement Rate Model

1.1. All-or-None Engagement Function

1.1.1. Value of the Informative Option

For the engagement function defined in Equation 3, let us assume the probability threshold for engagement is $\theta = .1$. Depending on the value of p, the animal may or may not engage in the Informative Option S_p trials.

When, in the Informative Option $\varepsilon(p) = 0$ (i.e., $p < .1$), the animal only engages with the S_1 stimulus (cf. Figure 2). Thus, to receive one unit of food the animal spends D + 1 units of time (with 1 corresponding to the handling time or the duration of food access). In this case, the reinforcement rate is given by

$$R_{Info\ \varepsilon=0} = \frac{1}{D+1} \tag{B1}$$

When, on the other hand, $\varepsilon(p) = 1$ in the Informative Option (i.e., $p \geq .1$), the reinforcement rate depends on the value of p (cf. Figure 2) and is given by.

$$R_{Info\ \varepsilon=1} = \frac{.2+.8\times p}{.2\times\left(D+1\right)+\ .8\times\left[\ p\times\left(D+1\right)+\left(1-p\right)\times D\right]} \tag{B2}$$

In this case, the animal receives food with probability .2 (E_1 trials) plus $.8\times p$ (E_p trials; see numerator of Equation B2). To obtain these rewards, with probability .2, the animal waits D+1 seconds (E_1 trials, with 1 as the handling time, h), with probability $.8\times p$ the animal waits D+1 seconds (E_p trials), and with probability $.8\times(1-p)$ the animal only waits D seconds for no reward (E_0 trials; see denominator of Equation B2).

1.1.2. Value of the Non-informative Option

Regarding the Non-informative Option, the animal engages in all trials ($\varepsilon(p) = 1$), and therefore the animal receives a reinforcer every two trials on average (see Figure 2). Hence, it has to spend about 2×D+1 seconds ($h = 1$) for each reinforcer. The reinforcement rate is

$$R_{Non-info\ \varepsilon=1} = \frac{1}{2D+1} \tag{B3}$$

1.2. Linear Engagement Function

1.2.1. Value of the Informative Option

If the engagement function is linear as defined in Equation 4, the reinforcement rate in the Informative Option depends on both p and ε(p), but because ε(p) = p we can substitute ε(p) by p and we obtain

$$R_{Info\ \varepsilon=p} = \frac{.2+.8p^2}{.2\times(D+1)+.8p\times\left[p\times(D+1)+(1-p)\times D\right]} \quad \text{(B4)}$$

which simplifies to

$$R_{Info\ \varepsilon=p} = \frac{.2+.8p^2}{.8p^2+.8pD+.2D+.2} \quad \text{(B5)}$$

In this case, the animal receives food with probability .2 (E_1 trials) plus 8×p×p (E_p trials; numerator of Equation B4). To obtain these rewards, with probability .2 the animal waits D+1 seconds (E_1 trials), with probability .8×p×p the animal waits D+1 seconds (E_p trials), and with probability .8×p×(1-p) the animal only waits D seconds (E_0 trials; denominator of Equation B4).

1.2.2. Value of the Non-informative Option

In the Non-informative Option, because the probability of reinforcement is .5 the animal only engages in half of the trials and thus only takes into account half of the reinforcers and the delays to obtain them. The rate of food intake in the Non-informative Option is given by

$$R_{Non-info\ \varepsilon=p} = \frac{.5\times.5}{.5\times.5\times(D+1)+.5\times.5\times D} = \frac{1}{(D+1)+D} \quad \text{(B6)}$$

The numerator indicates the probability of food: ε(p)×.5 = .5×.5 (E_1 trials). The denominator shows the time spent to obtain (and consume) the rewards: with probability ε(p)×.5 = .5×.5 the animal waits D+1 seconds (E_1 trials), and with probability ε(p)×.5 = .5×.5 the animal only waits D seconds (E_0 trials).

Thus, as in the Hyperbolic Discounting Model, the reinforcement rate when the engagement function is all-or-none or linear is the same:

$$R_{Non-info\ \varepsilon=p} = R_{Non-info\ \varepsilon=1} = \frac{1}{2D+1} \quad \text{(B7)}$$