Universidade do Minho
Escola de Psicologia

Álvaro Viúdez González

Rule-governed behavior and the description-experience gap

janeiro de 2018

Álvaro Viúdez González **Rule-governed behavior and the description-experience gap**

UMinho | 2018

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

POPH
QUALIFICAR É CRESCER.

QREN
QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
PORTUGAL 2007.2013

Governo da República Portuguesa

UNIÃO EUROPEIA
Fundo Social Europeu

Universidade do Minho
Escola de Psicologia

Álvaro Viúdez González

Rule-governed behavior and the description-experience gap

Tese de Doutoramento em Psicologia Básica

Trabalho efetuado sob a orientação do
Dr. José Keating
e da
Dra. Joana Arantes

janeiro de 2018

DECLARAÇÃO

Nome: Álvaro Viúdez González

Endereço eletrónico: alvaro.viudez@gmail.com **Telefone:** +34 687 062 644

Número do Bilhete de Identidade: 30250926-S

Título tese: Rule-governed behavior and the description-experience gap

Orientador(es): Dr. José Keating e Dra. Joana Arantes **Ano de conclusão:** 2018

Designação Doutoramento: Doutoramento em Psicologia Básica

Universidade do Minho, 25/01/2018

Álvaro Viúdez González

DECLARAÇÃO

Nome: Álvaro Viúdez González

Endereço eletrónico: alvaro.viudez@gmail.com **Telefone:** +34 687 062 644

Número do Bilhete de Identidade: 30250926-S

Título tese: Rule-governed behavior and the description-experience gap

Orientador(es): Dr. José Keating e Dra. Joana Arantes **Ano de conclusão:** 2018

Designação Doutoramento: Doutoramento em Psicologia Básica

Universidade do Minho, 25/01/2018

Álvaro Viúdez González

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração da presente dissertação. Confirmo que em todo o trabalho conducente à sua elaboração não recorri à prática de plágio ou a qualquer forma de falsificação de resultados.

Mais declaro que tomei conhecimento integral do Código de Conduta Ética da Universidade do Minho.

Universidade do Minho, 25/01/2018

Álvaro Viúdez González

- *At least I learned something from all of this…*
- *What's that?*
- *How to deal with frustration, disappointment, and irritating cynicism.*

Guybrush Threepwood, *The Secret of Monkey Island*^{TM}

A Santiago, por servirme como modelo y fuente de inspiración,

en la ciencia y en la vida.

Acknowledgments

I would like to thank my advisors, José Keating and Joana Arantes, for the great personal effort they have made to give me this opportunity.

The present work was supported by a Portuguese Foundation for Science and Technology (FCT) Doctoral Grant (SFRH/BD/52397/2013).

Rule-governed behavior and the description-experience gap

The present Ph.D. thesis assesses the relation between rule-governed or instruction following behavior and the description-experience gap from a behavioral perspective. Extensive research on the field of rule-governed behavior has shown that human behavior is often insensitive to schedule contingencies when instructions are available. On the other hand, the description-experience gap refers to the difference on choice behavior when either descriptions or experience are available. For our purpose, a meta-analysis of the existing literature on the description-experience gap was first conducted to explore interesting findings that could lead our work. Different factors were observed to influence the description-experience gap, namely: as-if over- and under-weighting functions of rare events in description- and experience-based tasks, task domain, probability of the rare event, existence of a certain option, description task paradigm and experience task paradigm. Besides obtaining these results, we found two studies on the description-experience gap domain showing absence of influence of the descriptions when personal experience is available. This intriguing result contradicts previous research on rule-governed behavior. In a series of three empirical studies, we subsequently evaluated the contribution of different moderator variables on the relation between rule-governed behavior or instructional control, and the description-experience gap. Specifically, we assessed the effect of the difficulty of the descriptions, the type of schedule, and the relative expected value between options. The absence of instructional control on studies previously conducted on the description-experience gap was found to not be caused by the difficulty of probabilistic descriptions or the type of schedule. However, the relative expected value between options demonstrated to modulate this effect. The relevance of this factor seems to rely on its capacity to help participants better discriminate the best option.

Comportamento governado por regras e a brecha descrição-experiência

A presente tese de doutoramento avalia a relação entre o comportamento governado por regras ou seguimento de instruções e a brecha descrição-experiência desde una perspetiva comportamental. Numerosas investigações no campo do comportamento governado por regras tem mostrado que o comportamento humano é frequentemente insensível às contingências programadas em presencia de instruções. Pelo outro lado, a brecha descrição-experiência refere à diferencia em comportamento de escolha em presença de descrições ou experiência. Para nosso propósito levou-se inicialmente a cabo um meta-análise da literatura existente sobre a brecha descrição-experiência para explorar resultados interessantes que pudessem guiar nosso trabalho. Diversos fatores demostraram possuir influência na brecha descrição-experiência, a saber: funções como-se se sobre- e sub-ponderassem os eventos raros em tarefas baseadas em descrições e em experiência, domínio da tarefa, probabilidade do evento raro, existência de uma opção segura, paradigma da tarefa de descrição e paradigma da tarefa de experiência. Ademais de obter estes resultados, encontramos dois estudos no domínio da brecha descrição-experiência que mostravam ausência de influência das descrições em presença de experiência pessoal. Este fascinante resultado contradisse investigações prévias em comportamento governado por regras ou controlo instrucional. Posteriormente, avaliámos numa série de três estudos empíricos a contribuição de diferentes variáveis moderadoras à relação entre comportamento governado por regras ou controlo instrucional, e a brecha descrição-experiência. Concretamente, avaliámos o efeito da dificuldade das descrições, o tipo de programa, e o valor esperado relativo entre opções. A ausência de controlo instrucional em estudos prévios sobre a brecha descrição-experiência revelou não estar causada pela dificuldade de descrições probabilísticas ou o tipo de programa. Contudo, o valor esperado relativo entre opções demostrou modular este efeito. A relevância de este fator parece dever-se à sua capacidade para ajudar aos participantes à discriminar com maior facilidade a melhor opção.

Table of Contents

Studies 3 and 4

Discussion

Abbreviations, acronyms and symbols

CR – Continuous Reinforcement

EV – Expected Value

FI – Fixed Interval

FR – Fixed Ratio

n – Sample size

p – Probability

SE – Standard Error

VI – Variable Interval

VR – Variable Ratio

w – Weight

Figures

Tables

Prologue

You wouldn't have a match… by any chance, would you?

This is how Chuck Noland (played by Tom Hanks) first addresses Wilson, a friend made out of a volleyball in the movie *Cast Away* (Zemeckis, 2000).

In this film, Tom Hanks plays the role of a FedEx executive forced to survive on a deserted island following a plane crash. The relationship that he establishes with Wilson has become iconic as an illustration of how socializing is as important for any human as getting potable water, finding food or making fire.

As humans, we socialize mainly through language. For this reason, language deserves to be scientifically investigated, so that we have a better understanding of this significant motor of our behavior. This Ph.D. thesis evaluates one of the functions that language can have with other behavior: rule-governed behavior or instructional control.

However, if Psychology is to help us guide our ship as we sail through the oceans of language, we should ask first:

What is Psychology?

Introduction

Opening remarks

A clear definition of the philosophical coordinates is of particular relevance on the present work in order to avoid possible misconceptions about the rational of the Ph.D. thesis itself. Therefore, we will start discussing some philosophical foundations before entering on the scientific process. Once this meta-theoretical framework is described, an introduction to choice behavior from a behavioral perspective will follow. Finally, rule-governed behavior or instructional control will be presented as the specific function to be evaluated on this Ph.D. thesis. This road from the general to the particular will find a symmetrical structure on the discussion, the studies serving as the central part of the work.

Towards a materialistic definition of Psychology

Natural and human sciences

To define Psychology as a science requires looking at it through the lens of Philosophy, as a knowledge of second grade that is capable of dealing with the ideas that emerge from the knowledge of first grade, that is, the sciences (Álvarez, 2008).

Spanish philosopher Gustavo Bueno is the main proponent of philosophical materialism (Bueno, 1972), a philosophical system that is of special relevance to define the philosophical foundations of a science. Within its coordinates, two dimensions are needed to differentiate natural sciences from human sciences: the distinctions between apothetic and parathetic relations and between alpha and beta situations.

Apothetic designates a relation that is given between two terms on the distance, the intermediate elements being dispensable for the analysis. It contrasts to parathetic relations, which are defined between objects that share contiguous relations (Bueno, 1995). The perception of an object is applies to the former group, while the physiological activity between the eye and the occipital cortex belong to the later.

The distinction between alpha and beta situations relies on whether the operatory subject (i.e., the idiosyncratic person) is formally included between the terms of the science or not (Bueno, 1995). Analyzing the factors that lead a child to cry so his parents buy him a toy would be delimited in an beta situation. On the contrary, the research field of the generic bodily process of crying, common to every person, refers to a alpha situation.

These dimensions are a powerful tool to categorize natural and human sciences. While parathetic relations and alpha situations are characteristic of natural sciences, the study of apothetic relations and beta situations are idiosyncratic of human sciences. Consequently, following this systematic classification, Psychology would be categorized as a human science. As a science which formally includes the operatory subject and interested in the study of the apothetic relations which these subjects establish with their context.

The science of Psychology

Psychology being classified as a human science is just the first step to define it, as there are other sciences included on this category. Philosophical materialism proposes the establishing of categorical closures to differentiate the sciences between them.

A categorical closure refers to the process of construction in which certain objects that are related to each other are composed or decomposed between them by certain operations resulting in third objects that still belong to the same gender of the former objects. In Chemistry, when two elements of the periodic table are combined using a particular chemical bond, the resulting molecule still belongs to the gender of chemical elements, and therefore to the categorical closure of Chemistry (Bueno, 1995).

By means of this definition, an important property arises: the irreducibility of the sciences between them. Each phenomenon is to be explained using elements and relations from the same science. Communication between different sciences is not denied from this perspective, but their delimitation is crucial to avoid fundamental mistakes. In a conversation with Nobel Prize winner Severo Ochoa, who was declaring that everything is Chemistry, Gustavo Bueno asked whether the characters of a book he was holding were attached by whether ionic or covalent bonds (Méndez, 2014). With this apparently naïve question, the Spanish philosopher was actually arguing that Grammar cannot be reduced to Chemistry, as the assertion of Ochoa implied.

However, there is no unified Psychology at all, but a multiplicity of them. Therefore, in order to maintain a categorical closure during the scientific itinerary of the present work, the paradigm of reference shall be defined. For this purpose, the basic foundations of radical behaviorism established by Burrhus Frederic Skinner will follow.

Radical behaviorism

Behavior has that kind of complexity or intricacy which discourages simple description and in which magical explanatory concepts flourish abundantly. Primitive systems of behavior first set the pattern by placing the behavior of man under the direction of entities beyond man himself.
(Skinner, 1938, p. 3)

In these words, Skinner establishes the main difference with other Psychology paradigms: the study of behavior for its own sake. Skinner proposes the behavior to be first described and then studied as a function of the variables of the context, taking into account that the individual himself and his history are also part of his context.

This definition of Psychology is clearly contrary to the study of the human mind and its mechanisms through the study of behavior, as is the situation in most of the contemporary Psychology. Appealing to concepts as “self”, “long-term memory”, “central executive” or “ego” as the causes of behavior, is committing the fallacy of begging the question because of their dualistic conception of human behavior: these mental mechanisms are supposed to explain overt behavior, but their existence can only be inferred through that same behavior.

From a behavioristic point of view, another common mistake that can be encountered in cognitive neurosciences is the reductionist monism. When an attempt to explain the behavior as being caused by its concomitant neural activity is made, researchers fall in the category mistake (Ryle, 1949). On this reductionist explanation, a human activity (e.g., trusting someone) is reduced to one of its elements (e.g., oxytocin levels); disregarding that the physiological activity is a part of what was to be explained on the first place, and also that Psychology is non-reducible to Biology.

Therefore, behaviorism’s goal is to explain behavior by appealing to the interrelations between the subject and the context, both present and past. Skinner based his experimental approach to complex human behavior such as emotion, self-control or thinking (Skinner, 1953), on Thorndike’s law of effect.

Choice on Experimental Analysis of Behavior

Law of effect

Before his famous experiments with cats, Edward Lee Thorndike conducted some previous research in animal intelligence during his stay in Harvard in 1897 with chicks. These experiments were conducted in William James' personal house basement (Tortosa & Civera, 2006, p. 258) and would lead Skinner some years later to establish the foundations of operant conditioning. Using stacked books as walls, Thorndike constructed different mazes, where he would put individual subjects and measure the evolution of time that was needed to successfully solve the maze. Thorndike found that this time was a decreasing function of experience (Thorndike, 1911, Figure 18): the animals were learning.

When Thorndike moved to Columbia on that same year to start his Ph.D., James McKeen Cattell got him a space in Columbia where Thorndike could establish an animal lab and conduct his historical experiments with cats (Hothersall, 1997, p. 386). Thorndike built different escape boxes where individual cats were locked in. In order to get out of the box, the animals needed to carry out a particular action in each box, such as pressing a lever, pulling a loop or turning a button. Again, the time required by the subjects to solve the box decreased over trials (Thorndike, 1898).

These experimental results would lead Thorndike to enunciate some years later his famous law of effect: "any act which in a given situation produces satisfaction becomes associated with that situation, so that when the situation recurs the act is more likely than before to recur also." (Thorndike, 1905, p. 203)

Thorndike's law of effect can be interpreted as sort of Darwin's theory of evolution acting at the ontogenetic level, by the mechanisms of variation and selection (Darwin, 1859). At a given situation where the organism is struggling for life, as when it is hungry, different behaviors will be exhibited –variation–. However, only some of them will eventually get the individual to solve the problem, making these responses more adaptive than the rest of them for these situations, and therefore more frequent –selection–.

Operant conditioning

The law of effect was re-elaborated and introduced under a behaviorist approach with the definition of operant behavior (Skinner, 1938, p. 19). The term operant behavior refers to the majority of adult, human behavior, which is learned as a function of the consequences that follow behavior. These consequences are functionally defined as reinforcers, and there are two types of them: positive and negative. Positive reinforcers are those which following a certain behavior increase its probability, as negative reinforcers have the opposite effect. While the termination of a positive reinforcer functions as a negative reinforcer, the termination of a negative reinforcer functions as a positive reinforcer (Skinner, 1938, p. 66).

Sometimes, each response that the organism makes under a reinforcement schedule is followed by the reinforcers. However, this is not always the case. Most of our behavior is only followed by reinforcers occasionally. These procedures are named continuous reinforcement, and intermittent or partial reinforcement, respectively (Skinner, 1938, Chapter 4).

There are four main schedules of partial reinforcement depending on which response will be followed by the reinforcer: fixed-ratio (FR), variable-ratio (VR), fixed-interval (FI) and variable-interval (VI) (Ferster & Skinner, 1957). Under ratio schedules, once a certain number of responses are made, reinforcer will follow. On the other hand, interval schedules are defined by a certain amount of time; once this time has elapsed, the first response will result in reinforcement. The difference between fixed and variable schedules relies on whether the parameter defining the schedule (i.e., number of responses on ratio schedules, and amount of time on interval schedules) is either constant or variable through different trials.

Stimulus discrimination

When an organism is discriminating one stimulus from another, that is, behaving differently in the presence of one or the other, three basic elements are needed for consideration (Skinner, 1938, p. 169). On the first place, the stimulus whose presence signals the occasion for emitting a certain behavior, called discriminative stimulus. The next element is the response itself, and finally, the reinforcer. On the other side, there shall be at least one stimulus whose presence signals the occasion for not emitting that behavior, called delta stimulus. This is, in the presence of the delta stimulus, if the aforementioned response is emitted, no reinforcer –or a negative one– will appear. While a reinforcement schedule is operating on the discriminative stimulus, an extinction or punishment schedule is operating on the delta stimulus. With this procedure, the subject will eventually start discriminating the discriminative from the delta stimulus, by responding almost exclusively on the former while virtually never on the later.

Concurrent schedules

When the schedules of reinforcement and stimulus discrimination are combined, the fundamental method of measuring choice behavior under a behaviorist approach is obtained: concurrent schedules. With this method, two discriminative stimuli signal two different schedules of reinforcement, and the subject can distribute behavior between both (i.e., two responses collectors are available simultaneously). It was first developed by Ferster and Skinner (1957) and has become a main instrument for measuring choice behavior in experimental analysis of behavior since then.

One of the most powerful tools that has been developed on this field refers to the concurrent variable-interval schedules. On these procedures, each component will reinforce the first response that is made once the established time for that schedule has elapsed. For example, in a concurrent VI 60-s VI 30-s schedule, the first component will deliver a reinforcer when the individual makes a response on its response collector after an average time of 60 s has elapsed. Once the reinforcer is delivered, the schedule will continue as usual. On the second component, the same rules apply, being that it will deliver reinforcers with double frequency (i.e., after an average time of 30 s).

The matching law was the first formal expression of the behavior that is exhibited under this type of schedule. It states that subjects match relative frequencies of responses in both schedules to the relative frequencies of reinforcements (Herrnstein, 1961):

$$\frac{B_1}{B_1 + B_2} = \frac{R_1}{R_1 + R_2}$$

Or a mathematically identical version,

$$\frac{B_1}{B_2} = \frac{R_1}{R_2}$$

Where *B* denotes behavior and *R* denotes number of reinforcers. However, this equation makes it difficult to interpret results of subjects that do not exhibit perfect matching. The generalized matching law (Baum, 1974) is an useful tool to account for deviations from matching:

$$\left(\frac{B_1}{B_2}\right) = b\left(\frac{R_1}{R_2}\right)^s$$

As can be mathematically derived, the original and the generalized matching law are identical when *b* and *s* equal 1. On this equation *b* denote a systematic response bias towards one component. Values of *b* higher and lower than 1 represent biases towards component 1 and 2, respectively. On the other hand, *s* represents sensitivity to changes in relative frequencies of reinforcement. Values of *s* higher and lower than 1 denote over- and under-sensitivity: the subject changes the relative response frequencies at a higher or lower rate than the changes on relative reinforcement frequencies, which can be understood as over-matching or under-matching, respectively.

Also, a subsequent logarithmic transformation was constructed to facilitate a linear interpretation of results (Baum, 1974):

$$\log\left(\frac{B_1}{B_2}\right) = s \log\left(\frac{R_1}{R_2}\right) + \log b$$

By using the logarithmic version of the generalized matching law, the polynomial function is transformed to a linear equation where deviations from perfect matching can be easily interpreted. Systematic deviations on the intercept of the function represent bias towards one option, while systematic deviations on the slope of the function represent sensitivity biases (see Figure 1). This approximation to choice behavior will be a critical part in Study 4.

The different versions of the matching law have proven to be an useful approach on natural human situations such as scholar settings (Neef, Mace, Shea, & Shade, 1992; Neef, Shade, & Miller, 1994), disruptive behavior (Martens & Houk, 1989), self-injurious behavior (McDowell, 1988), and developmental disabilities (Athens & Vollmer, 2010), among others.

However, our social nature gives us the opportunity to modify our behavior through language so we do not need to learn everything by having direct, naïve experience. When conducting scientific research, there is always some kind of instructions for the participants to behave in a certain manner, even if some information is hidden from them. The differential influence of both experimental instructions and schedule contingences is assessed on the research field of rule-governed behavior or instructional control.

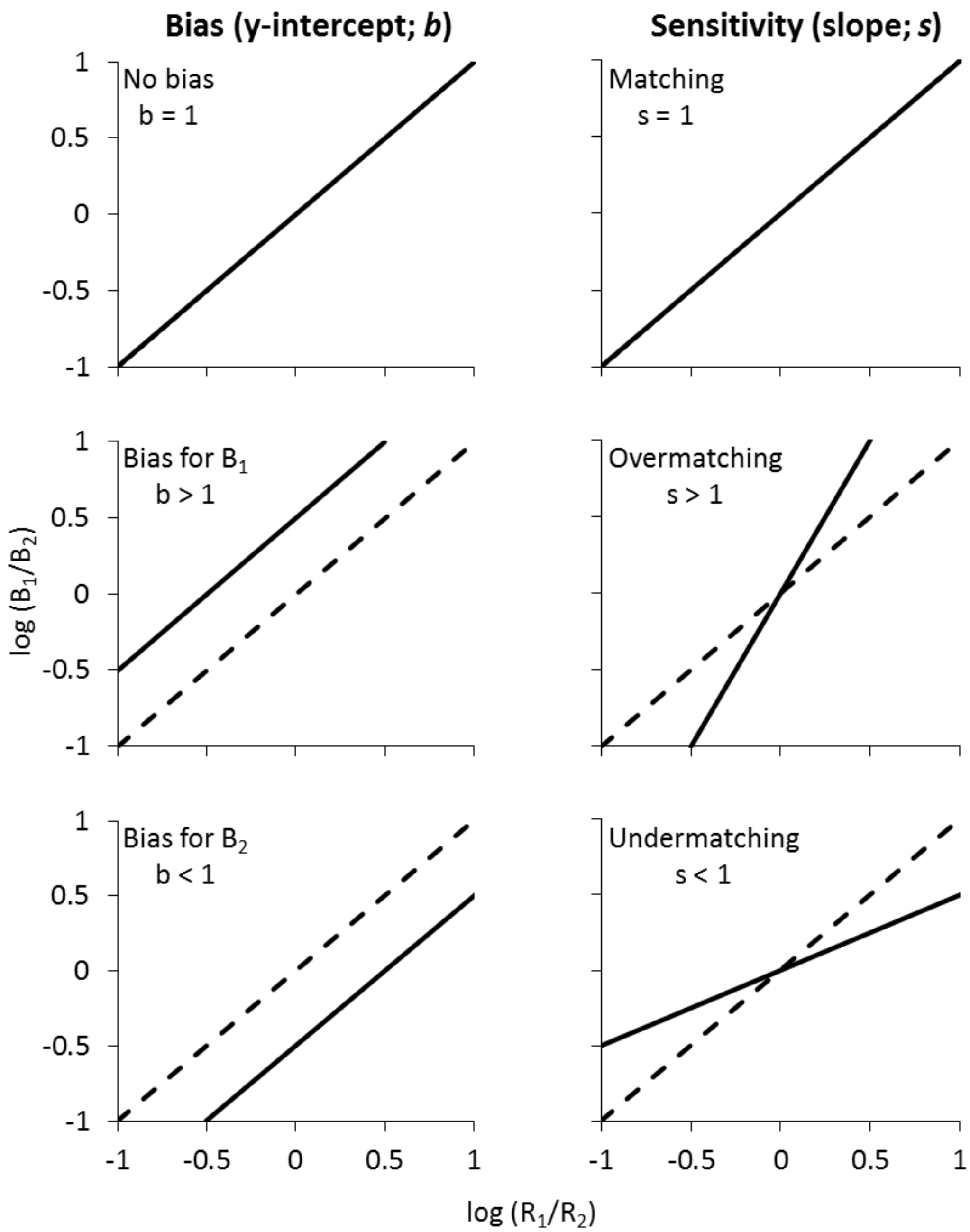

Figure 1. Systematic deviations from matching represented as logarithm transformations of responses rates against reinforcers rates. Left and right panels show deviations in bias and sensitivity parameters, respectively. Adapted from "The matching law: a tutorial for practitioners," by D. Reed and B. Kaplan, 2011, *Behavior Analysis in Practice, 4*, 2, p. 15. Copyright 2011 by ABA International.

Instructional control

Contingency-governed and rule-governed behavior

Bertrand Russell (1910) distinguished between knowledge by acquaintance and by description: while the former is given when the subject directly perceives the object to be known, the later results from linguistic descriptions of it. Some years later, famous philosopher Gilbert Ryle dedicated a whole chapter on his book *The Concept of Mind* to what he called knowing how and knowing that. When a person knows how to do something, he is apt to meet some criteria, adapt his behavior to eventual changes, and improve on his execution. On the other hand, when a person knows that something is, he is able to describe the characteristics that are of concern (Ryle, 1949).

In order to maintain a categorical closure during the present work, these distinctions can be covered and interpreted under Skinner's behaviorist proposal of contingency-governed and rule-governed behavior. On *Verbal Behavior,* Skinner (1957) started a research field on Psychology of Language without falling in a dualist conception of the subject. However, this proposal was focused on creating a functional taxonomy of the verbal behavior of the speaker rather than the listener. The distinction between contingency-governed and rule-governed behavior was first explicitly established on *Contingencies of Reinforcement: A Theoretical Analysis* (Skinner, 1969): the former is shaped by the reinforcement contingencies, while the latter is shaped by the description of those.

Functional definition of rule

Skinner (1969) functionally defines a rule as a discriminative stimulus; one that the subject creates describing the contingencies of the behavior of interest. Therefore, it signals the occasion in which a particular behavior will be reinforced, as every other discriminative stimulus. However, it is to be noted that it has a linguistic nature. It is because of this nature that the person is able to disentangle his behavior from its actual contingencies and anticipate to situations or elaborate plans. Human activities such as chemistry, driving or playing an instrument would need a much longer time to be learned, and the acquired proficiency would be much lower, if no information about the periodic table, traffic signals, or musical notes were given.

According to Ribes-Iñesta (2000), these verbal descriptions of the contingencies should be considered functionally different depending on whether the subject creates them after having been exposed to the contingencies or they are given to the subject without direct experience with the contingencies. The former are considered rules while the later are considered instructions. As the present work is interested on the later, the next section will show different examples of instructional control; the control that instructions can exert over behavior.

Experimental precedents of instructional control

The presence or absence of instructions about the operating schedule and appropriate response can yield several consequences upon participants' behavior (Hayes, 1989). Different empirical results showing these effects will be discussed on the following paragraphs.

The absence of instructions may make it difficult for appropriate responding to establish. This effect was tested by Ader and Tatum (1961) in a free-operant avoidance conditioning to examine whether human participants would acquire appropriate avoidance behavior –pressing a button to avoid electric shocks– without giving instructions about the response and its relation with the contingencies. More than half of the participants failed to acquire appropriate behavior. Similar results were encountered by Turner and Solomon (1962) using different responses' topographies. Of four different responses that were studied, three showed poor or no acquisition of proper avoidance behavior when no instructions were given. On the other side, almost every participant receiving appropriate instructions learned to avoid shocks, which leads us to the next point.

Appropriate instructions can facilitate rapid acquisition of appropriate response. Ayllon and Azrin (1964) studied the effect of instructions on 18 patients of a mental hospital to establish socially acceptable eating behavior of picking up all of the cutlery on their meals. On the first phase of the experiment, they would be reinforced with their choice of a piece of candy, a cigarette, an extra cup of coffee, or an extra glass of milk, when an appropriate response was made. After 20 meals comprising only reinforcement contingencies, detailed instructions were given. While the initial contingencies made no effect on establishing the behavior, once the patients received instructions the proportion of participants picking up all the cutlery raised from 10% to 66%. Half of the participants that reached the criterion continued emitting appropriate behavior one year after the experiment stopped. On a second experiment, a different group of patients were given only instructions on a first phase lasting 110 meals and both instructions and consequences on a second phase. The proportion of patients raised from 25% to 95% when both instructions and consequences were arranged (Ayllon & Azrin, 1964). Therefore, instructions can induce in some subjects appropriate responding even in the absence of scheduled consequences. However, they are more effective when both are functioning. Baron and Kaufman (1966) found similar results using a free-operant avoidance schedule with monetary reinforcement. Participants would earn money at different intervals unless a signaled time out was in effect. Time out periods could be escaped or avoided by pressing a button. While on the first session including the time outs almost no response was given, appropriate responding was established in three of four participants almost immediately when they were told that something could be made to avoid time outs.

Participants' response rates in different partial reinforcement schedules are an interesting way of assessing appropriate behavior. There are divergent results regarding appropriate response rates when no specific instructions are given to the participants. For example, Holland (1958) studied observing responses on a signals detection task with Navy enlisted men. Working in a dark room, participants had to report deflections of a pointer on a dial. Due to the lack of illumination, the dial could only be seen by pressing a button that made a light illuminate it by flashing during .07 s. Dial deflections happened by following different fixed-interval, fixed-ratio and differential-reinforcement-of-low-rate (i.e., two responses separated in time by a certain amount of seconds) schedules on this button. Participants' response rates differentiated through the different schedules and resembled those of non-human subjects. Also, Matthews, Shimoff, Catania, and Sagvolden (1977), by shaping key pressing instead of instructing the participants as usual, found differences in key responding between variable-ratio and variable-interval schedules similar to those observed in non-human animals: a high and steady response rate on the variable-ratio schedule, while a much lower rate on the variable-interval schedule. This result was observed both in a between- and a within-subjects design.

In contrast with these results showing appropriate adaptation of behavior to different schedule contingencies, some studies found the opposite effect. Blair (1958) used a similar procedure to that of Holland (1958), but changed the nature of the observing response from button presses to orientation movement of the participant's head, where a continuous light source was fixated, towards the dial. Participants worked on 30 min sessions using a FI 60-s schedule. Two of the five participants exhibited typical post-reinforcement pauses, while three did not. Possible causes of this difference include the nature of the response, or the differences between participants (Blair, 1958). Another example can be found on the experiment of Weiner (1962), asking participants to report a red signal when it was active to earn points. In order to see whether it was active or not, they had to press a button that would turn on the red signal when it was active, and remain black when it was not. Fixed-interval and variable-interval schedules were used. Participants displayed similar, high rates across the different variable-interval schedules until a response cost was introduced in the experiment. Finally, Galizio (1979) conducted a series of experiments using a free-operant avoidance schedule. Participants had to operate a lever to avoid point loss under a multiple schedule including different variable-interval schedules and one extinction schedule (i.e., no point loss was programed). On a multiple schedule, each component is signalized by a discriminative stimulus and remains active for a certain period of time before switching to the next one. On the first condition of Experiment 1, participants received no instructions about the different schedules. Of four participants, three did not differentiate their response rates across conditions, resulting in high response rates even when the extinction schedule was operating. Compared to the aforementioned results, these three experiments showed that participants' behavior were sensitive to schedule contingencies only when either the behavior itself or the contingencies associated to it became more demanding.

However, once detailed instructions about schedule contingencies and responses are given, they acquire control over participants' behavior. Under this condition, participants from an experiment of Dews and Morse (1958) behaved efficiently in a combined fixed-ratio and differential-reinforcement-of-low-rate schedule. In such a schedule, participants have to respond a certain number of times (i.e., fixed-ratio), being their responses separated in time by a certain amount of seconds (i.e., differential-reinforcement-of-low-rate). This result was observed both in a control condition and in a drug condition where participants were given dextro-amphetamine. Weiner (1962) obtained similar results studying the effects of response cost upon observing behavior using different variable-interval schedules. Participants receiving information about the schedules showed different rates across conditions as opposed to those who did not. Furthermore, Baron, Kaufman, and Stauber (1969) used a five-component multiple schedule composed of different fixed-interval schedules. On this experiment, each component lasted 10 min and the fixed-interval values were 10, 30, 90 and 270 s. The last component was an extinction schedule. The order of the components were pseudo-random. The authors found that participants that had instructions about the schedule functioning exhibited more appropriate and differentiated response rates between components compared to those participants without instructions. In a subsequent experiment, giving detailed instructions to some participants of the no-instructions group of the first experiment barely changed their behavior. Similar results were obtained by Galizio (1979).

One of the most interesting effects of instructional control is producing behavior which is not in accordance with schedule contingencies. Lippman and Meyer (1967) used a FI 20-s schedule to assess response rates on three different conditions: interval instructions, ratio instructions and no instructions. Participants on the first and second groups were instructed that the occasion to get points by pressing a button would appear after a certain amount of time after each point or after a certain number of button presses, respectively. Participants on the third group were not instructed about the contingencies at all. While the no instructed participants exhibited high variability on the response rates, the interval instructed participants showed a similar, scalloped fixed-interval pattern of responding. However, two out of three participants of the ratio instructed participants showed a steady, high rate of responding, typical of variable-ratio schedules. The third participant was the only one in this group that explicitly mentioned that she discovered that the occasion to get points did not depend on the number of responses but on the interresponse time, and she behaved similarly to the participants from the interval instructions group.

Additionally, second experiment of Galizio (1979) using a free-operant avoidance schedule instructed participants as if they were still participating on the multiple schedule with different variable-interval schedules and one extinction schedule, while only the extinction schedule was in effect (i.e., no point loss was programmed at all). Even if participants did not have to respond, they exhibited differential response rates as the instructions stated. Instruction following behavior stopped only when it produced negative consequences on a subsequent condition.

The enlisted results together with several other studies (Catania, Matthews, & Shimoff, 1982; Hackenberg & Joker, 1994; Matthews, Catania, & Shimoff, 1985; Matthews et al., 1977; Shimoff, Catania, & Matthews, 1981) demonstrate the powerful nature of instructions, as these are even capable of making participants' behavior insensitive to schedule contingencies. The relevance of this phenomenon comes because it will be contradicted by two studies that we found on the meta-analysis on the description-experience gap field of Study 1, as will be introduced on the following section.

Aims of the present work

The main goal of the present Ph.D. thesis is to assess the existence of rule-governed or instruction following behavior on the description-experience gap field. Extensive research on rule-governed behavior (see previous section) has shown that human behavior is often insensitive to schedule contingencies when instructions or descriptions are available. On the other hand, the description-experience gap refers to the difference on choice behavior when either descriptions or experience are available.

For our purpose, a meta-analysis of the existing literature on the description-experience gap was first conducted to explore interesting findings that could lead our work. Two interesting studies including a group having both description and experience available (i.e., instructions and schedule contingencies) served our purpose: these studies show absence of instructional control (Jessup, Bishara, & Busemeyer, 2008; Lejarraga & Gonzalez, 2011). The objective of the empirical experiments is to disentangle the factors that are causing this absence of instructional control. More specifically:

- The meta-analysis from Study 1 will evaluate the general framework of the description-experience gap in terms of the results and explanations that characterize the existing literature.
- The objective of Study 2 is to assess the probabilistic descriptions of the options as a plausible cause of the absence of their control on participants' behavior.
- Studies 3 and 4 will jointly evaluate the contribution of the type of schedule, and the relative expected values between options, to the existence of instructional control.

Study 1

The description-experience gap: a meta-analysis[1]

Choice behavior differs depending on how the information about the options is presented to the subjects, via descriptions or direct experience; a phenomenon called the description-experience gap. Cumulative Prospect Theory implies that subjects choose as-if they were overweighting rare events in description-based tasks. However, in experience-based tasks, the opposite result is found: subjects choose as-if they were underweighting rare events. The present meta-analysis studied three important factors on the description-experience gap related to Cumulative Prospect Theory: the as-if over- and under-weighting functions of rare events in description- and experience-based tasks, the task domain and the probability of the rare event. Asides from these three elements, another three additional factors were studied: the existence of a certain option, the description task paradigm and the experience task paradigm. All of them were found to be statistically significant on having a differential effect on the description-experience gap. We suggest that the fact that the reference model Cumulative Prospect Theory (Tversky & Kahneman, 1992) is a descriptive one, and the most appealed explaining factor is sampling bias –which is part of the methodology of the task itself– may be suggesting that the description-experience gap is an irreducible psychological phenomenon (i.e., a phenomenon that does not rely on other psychological mechanisms, but solely on the methodology of the task).

[1]This chapter reproduces the document: Viudez, A., Keating, J., & Arantes, J. (2018). The description-experience gap: a meta-analysis. *Manuscript submitted for publication.* Minor modifications were made to avoid overlap between different chapters.

Introduction

When confronted with formally expressed probabilities, we tend to choose as-if we overweight low probabilities and underweight high probabilities, as described by Cumulative Prospect theory (Tversky & Kahneman, 1992). Nonetheless, the opposite result is found when we have to experience the probabilities of the events ourselves rather than reading them, giving its origin to a phenomenon called the description-experience gap (Barron & Erev, 2003; Hertwig, Barron, Weber, & Erev, 2004). However, some studies argue that there is no description-experience gap, as it may be an artifact derived from some properties of the task, as biased samples (Camilleri & Newell, 2009b, 2011a; Glöckner, Hilbig, Henninger, & Fiedler, 2016; Rakow, Demes, & Newell, 2008). The present meta-analysis aims to evaluate the existing data on this literature.

Description vs. experience

The distinction between decisions from description and decisions from experience was introduced long time ago by Frank Hyneman Knight. He distinguished between decisions under risk, where the subject knows the *a priori* probabilities of different gambles, and decisions under uncertainty, where the *a priori* probabilities remain unknown (Knight, 1921).

In tasks involving description-based choices, different gambles are shown to the individuals in a text and/or graphic way (see Weber, Shafir, and Blais (2004) for a meta-analysis). Participants have all the information available from the beginning, that is, the outcomes values and probabilities are known.

In tasks involving experience-based choices, two options are shown to the participants in a symbolic way (e.g., doors, bags, buttons, etc.) (for a review of recent research, see Rakow and Newell (2010)). Therefore, they know nothing about the outcomes values and probabilities, being only able to infer them by sampling, with or without real consequences, from both options.

The description-experience gap

Barron and Erev (2003) published the first paper which explicitly compares on the same publication description-based with experience-based choices. These authors made ten comparisons between description-based and experience-based choices and found statistically significant differences in four of them: choice behavior was demonstrated to be different depending on how the information is achieved – through description or experience. The group facing experience-based choices "was found to lead to a reversed common ratio/certainty effect, more risk seeking in the gain than in the loss domain, and to an underweighting of small probabilities" (Barron & Erev, 2003). Decision weights (Tversky & Kahneman, 1992) are calculated to summarize already obtained results and to give a measure of the impact of each outcome (Hertwig et al., 2004). This is, when a subject chooses a risky, favorable option over a certain option with equal or higher expected value, it is assumed that he is choosing "as-if" that favorable outcome was being overweighted. These functions are graphically represented in Figure 2.

Following Barron and Erev's research, other studies have investigated the description-experience gap, both in basic tasks involving points/money and in applied tasks such as social cooperation (Martin, Gonzalez, Juvina, & Lebiere, 2014), online product reviews (Wulff, Hills, & Hertwig, 2015), climate change (Dutt & Gonzalez, 2012a, 2012b) or medical decisions (Lejarraga, Pachur, Frey, & Hertwig, 2016).

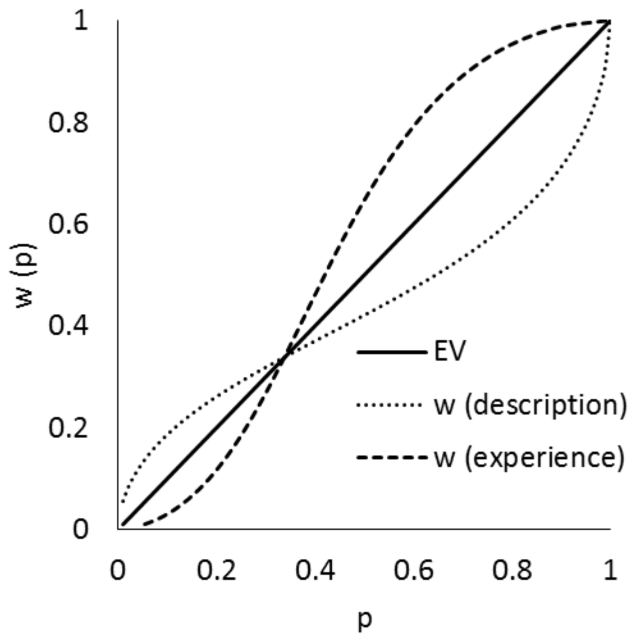

Figure 2. Graphical representation of the different proposed as-if weighting functions for description- and experience-based tasks.

The present work may help other researchers on the field to know the findings up-to-date of the description-experience gap and the explaining factors used by the respective authors, as there are only some non-systematic reviews on the field using a smaller number of publications (Erev & Roth, 2014; Hau, Pleskac, & Hertwig, 2010; Hertwig, 2012; Rakow & Newell, 2010). However, a meta-analysis that was being conducted parallel in time to the present one (Wulff, Canseco, & Hertwig, 2017) already tested some of the factors that we were working on. Even if a different measurement unit was used on the present meta-analysis –standardized mean differences– we adapted our analysis to report non-redundant information.

The other meta-analysis focuses on one type of experience task paradigm: the sampling paradigm (Wulff et al., 2017). Therefore, regarding the factors that overlap through both meta-analysis, we focused on the other experience task paradigm –the feedback paradigm– and also looked for differences at the aggregate level including both paradigms. These factors include:

1. **As-if weighting functions.** The assertion that rare events affect choice as-if they were over-weighted in description-based tasks while under-weighted in experience-based tasks (Barron & Erev, 2003; Hertwig et al., 2004) is a fundamental one in the description-experience gap field. This factor was also analyzed on the meta-analysis conducted on the sampling paradigm, but the authors did not conclude that different probability weighting is found between conditions (Wulff et al., 2017).
2. **Domain.** Cumulative Prospect Theory (Tversky & Kahneman, 1992) predicts a higher as-if overweighting of rare events for negative, compared to positive, outcomes. Therefore, a higher description-experience gap should be found on the former. This factor found favorable evidence on the meta-analysis conducted on the sampling paradigm (Wulff et al., 2017).
3. **Probability of the rare event.** Cumulative Prospect Theory predicts a higher deviation from objective probability near the extremes of the probability continuum (Tversky & Kahneman, 1992). Also, some authors have argued that the presence of rare events is a key factor in the description-experience gap (Hau et al., 2010; Rakow & Rahim, 2010). Therefore, their probability should reveal to have a differential impact: the lower the probability of the rare event, the bigger the description-experience gap. This factor found favorable evidence on the meta-analysis conducted on the sampling paradigm (Wulff et al., 2017).
4. **Existence of a certain option.** We could extend the same logic of the previous point to the present one: the difference between an event with a probability of .99 and a certain event is not just quantitative (i.e., a difference of .01 in probability) but also qualitative (i.e., a certain degree of ambiguity versus certainty). Therefore, we should find a difference between problems using one or the other. This factor found favorable evidence on the meta-analysis conducted on the sampling paradigm (Wulff et al., 2017).

Asides from these factors included in both meta-analysis, two additional points were exclusively studied on the present meta-analysis:

5. **Description task paradigm.** Some authors have argued that the likelihood of an event is harder to understand when using probabilities compared to frequencies or other methods (Gottlieb, Weiss, & Chapman, 2007; Harman & Gonzalez, 2015; Hilbig & Glöckner, 2011). Therefore, the description-experience gap should be greater when probabilities are used, as the as-if weighting functions of the events would lead to a greater difference from its objective odds.
6. **Experience task paradigm.** For the sole effect of reinforcement (Skinner, 1938), a more important consequence for the subject will have a higher impact on his behavior. Therefore, when a subject receives real feedback when choosing an option, his behavior will be more affected by those consequences compared to a situation without real feedback: we would expect a higher description-experience gap in the former. This factor was studied on the existing meta-analysis (Wulff et al., 2017) with a limited number of studies and including only one the partial-feedback paradigm, so we deemed appropriate to include at a higher extent a bigger picture on the present work. Actually, this is a crucial analysis, as it explicitly compares the paradigms studied individually by both meta-analysis (i.e., sampling and feedback).

Method

Inclusion/exclusion criteria

a. Type of participants

Studies with human subjects, no matter the age, were considered for this review.

b. Type of studies

Empirical studies on the description-experience gap were eligible for inclusion in this review. The description-experience gap had to be tested using studies from the same paper. Comparisons with data from other papers and data reinterpretations were excluded from this review.

This review includes studies that used the same or set of problems structures for different subjects in order to control for standard deviations. Therefore, studies using a different problem for each participant were excluded from this review.

In addition, participants had to experience directly the tasks. Studies that used the intermediation of a second participant between the subject and the outcomes were not eligible for inclusion in this review.

c. Type of task

In order to have compatible data for the meta-analysis, our criteria was restricted to studies using gambles for points/money. Studies using other types of choices, such as medical decisions, climate change decisions, etc., were excluded from this review. We do not deem appropriate to compare the proportion of people choosing a risky medical treatment versus a lottery for money.

The decision tasks eligible for inclusion had to ask the participant to choose between two or more options. When a in a particular study (Yoon, Vo, & Venkatraman, 2017) participants were asked to choose between three options, we calculated the description-experience gap as a function of the proportion of choices in the option containing the most extreme outcomes. The few studies that we found that used tasks that asked the participant how much would he pay for playing a certain option were excluded for inclusion in this review, as the data cannot be structured the same way than the others in order to be compared.

Primary outcomes

The average and standard deviation of the proportion of choices on the option containing the rare event was calculated for each description and experience group. Then, the description-experience gap was calculated by subtracting that proportion on the description group from the same proportion of the experience group (Glöckner, Fiedler, Hochman, Ayal, & Hilbig, 2012). Finally, standardized mean differences (i.e., Cohen's *d*) were computed for each comparison in order to have a better between-studies measurement unit.

Secondary outcomes

Asides from the aforementioned quantitative dependent variables, the factors that were considered in the original publications to explain the description-experience gap were enlisted for a qualitative assessment of the phenomenon.

Search strategy to identify studies

Electronic search was conducted using as keywords on the topic "description-experience gap", "decisions from experience" and "decisions from description" in *Web of Science* (http://ipscience.thomsonreuters.com/product/web-of-science/); and on the title and abstract in *Ovid*, including *PsycINFO*, *PsychArticles* and its own database (http://ovidsp.ovid.com/). Search results are updated up to August 10th, 2016.

Additional search on *Judgment and Decision Making* journal, *Journal of Behavioral Decision Making* and *Google Scholar* was conducted in order to get publications that were not found in the aforementioned databases and update the results up to May 15th, 2017.

Systematic review management

a. Selection of studies

Two reviewers (AV and JK) examined the abstract of every publication found by our search method. Once the potential publications were identified, they were explored to select those that met our criteria:

1. Empirical studies on the description-experience gap.
2. Participants were human subjects.
3. The tasks referred gambles for points/money.

Documents such as commentaries, notes, revision, indexes, editorials, erratum and books were excluded, as well as literature reviews and case studies. Eligibility was established in two stages, based on the screening of the abstract and the entire manuscripts. In case of doubts about the inclusion of a publication, the third author (JA) was asked for her opinion.

b. Data extraction

A summary table was constructed containing: i) publication identification, ii) number of subjects, iii) experimental design, iv) description task paradigm, v) experience task paradigm, vi) number of problems used, vii) type of outcome, and viii) main factors used to explain the description-experience gap (see Table 1).

c. Multiple groups management

Some publications only have a description group and an experience group (Artinger, Fleischhut, Levati, & Stevens, 2012; Camilleri & Newell, 2009a, 2009b; Glöckner et al., 2012; Hertwig et al., 2004; Kellen, Pachur, & Hertwig, 2016; Kudryavtsev & Pavlodsky, 2012; Lejarraga, 2010; Madan, Ludvig, & Spetch, 2017; Weber et al., 2004). Being that the case, the description-experience gap was calculated by comparing them.

However, the other publications have more than one description and/or experience group. Two possibilities determined how we analyzed the data in each case: either the same modification was applied to create new description and experience groups, or they were created using different modifications.

If the same modification is applied to create a new pair of description and experience groups, the data from these "pairs" was compared. For example: consider a publication that has two description and two experience groups. The only difference between the two description groups is that the participants have to choose which option they prefer if they were to play it once (group 1) or one hundred times (group 2). The same thing happens with the experience groups: after sampling 40 times, they have to choose which option they prefer if they were to play it one time (group 3) or one hundred times (group 4). If this is the case, the description-experience gap is calculated between the pairs 1–3 and 2–4.

If different modifications are applied to create new description and experience based groups (or if just one of the groups is modified), all the possible combinations were compared. For example: consider a publication having two description groups and three experience groups. The difference between the two description groups is that one of them uses probabilities (group 1), while the other uses frequencies (group 2) to describe the outcomes. On the other hand, the difference between the three experience groups is that one of them uses the sampling paradigm (group 3), another one uses the feedback paradigm (group 4), and the last one uses the controlled feedback paradigm (group 5). If this is the case, the description-experience gap was calculated between the pairs 1–3, 1–4, 1–5, 2–3, 2–4 and 2–5.

d. Yoked groups

In some studies, a group of participants do not face the *a priori* probabilities, but the probabilities experienced by participants in the experience group. In these cases, each participant from the former *yoked* group either experience (experience-based task) or read (description-based task) the outcomes values and probabilities that a particular subject of the experience group faced. So each participant on this *yoked* group has a “twin” in the experience group. This method is used to study the description-experience gap ensuring that participants in both groups face the same objective outcomes probabilities and sequences.

e. Single/multiple plays

Usually, the participants are asked about which option they would they like to choose, to play it once or repeated times. Nonetheless, in some studies they are asked about which option they would like to choose, for the computer to play it for them a certain number of times. Also, they could be asked about how they would like to distribute their choices between the two options if they had to play for a certain number of times.

Meta-analysis

The data analysis was conducted by following appropriate methods for subgroup analysis (Borenstein, Hedges, Higgins, & Rothstein, 2009; Borenstein & Higgins, 2013) using a custom made Microsoft Excel sheet for qualitative factors and Comprehensive Meta-Analysis software for the meta-regression. Meta-analyses require the standard deviation of each group to be different from zero, so when exceptional cases like this were encountered, an ad-hoc method was used by entering the average standard deviation of the corresponding groups.

A random-effects model within subgroups, fixed effects model between subgroups (Borenstein & Higgins, 2013) was used for our purposes. When appropriate, a Q or a Z-test for the difference in Cohen's *d* units was conducted, and a 95% confidence interval was obtained for this difference, together with the I^2 test for heterogeneity and the R^2 effect size. Six different factors were analyzed using this method:

a. As-if weighting functions of rare events

The rare event of a problem is defined by the outcome with the lowest associated probability. For example, in a problem containing an option A that gives 10$ with a probability of .20 and 5$ otherwise, and an option B that gives 8$ with a probability of .90 and 0$ otherwise, the rare event is 0$, whose probability is .10. If every outcome has a probability of .50, the problem has no rare event.

The rare event can be defined as either desirable or undesirable, depending on whether its outcome value is higher or lower, respectively, than the other outcomes values of the same option. On the example stated above, the rare event 0$ is undesirable, as 0$ < 8$.

Early studies on the description experience gap (Barron & Erev, 2003; Hertwig et al., 2004) stated that rare events are as-if overweighted in decisions from description while underweighted in decisions from experience (i.e., rare events have a higher impact in decisions from description). Therefore, when the rare event is desirable, this argument predicts a higher proportion of choices on that option in the description-based choices, and vice versa when the rare event is undesirable.

The description-experience gaps were calculated by subtracting the proportion of choices on the option containing the rare event on the description group from the same proportion of the experience group. Therefore, positive description-experience gaps mean higher proportion of choices on that option for the experience group, and vice versa for negative description-experience gaps. Being that the case, when the rare event is undesirable, we shall find a higher proportion of choices on that option on the experience group –positive description-experience gap–, being the opposite when the rare event is desirable –negative description-experience gap–.

b. Domain

The domain of the problem could be defined as either gain or loss, depending on whether every non-zero outcome is positive or negative, respectively. If a problem comprised both positive and negative non-zero outcomes, it implies a mixed domain. As Cumulative Prospect Theory predicts higher as-if overweighting function of negative rare events on the loss domain (Tversky & Kahneman, 1992), a greater description-experience gap should be found when compared to the gain domain.

c. Probability of the rare event

Cumulative Prospect Theory assigns different weights to rare events depending on their probabilities: "departures from linearity, which violate expected utility theory, are most pronounced near the edges" (Tversky & Kahneman, 1992). Therefore, we expect the description-experience gap to have a negative correlation with rare event probability.

Data from the third experiment from Glöckner, Hilbig, Henninger, and Fiedler (2016) was not included on this analysis, as their problems made the experienced probabilities fluctuate, while the list with the original problems was not available in order to know the *a priori probabilities.* Also, data from studies without rare event (i.e., every outcome having a probability of .5) were excluded from this analysis.

d. Existence of a certain option

Once we know how the option containing the rare event is defined, there are two possibilities for the "other option": whether it has one possible outcome, or more than one.

If it has more than one possible outcome it is called a *riskier* option, as its result is necessarily harder to predict compared to the option containing the rare event. This happens because, on the "other option", the outcome with the lowest probability still has a higher probability than the rare event, by definition. As every outcome probability on this "other option" is nearer to .50 than on the option containing the rare event, the "other option" is the one whose risk is higher.

If it has just one possible outcome, it is called a *certain* option. The uncertainty is null, and the option having the rare event becomes also the riskier option. For example, in a problem that has an option A that gives 10$ with a probability of .20 and 5$ otherwise, and an option B that always gives 7$, the option A contains the rare event 10$, and also becomes the riskier option.

A binary factor was used to test the magnitude of the description-experience gap depending on whether the other option has one or more than one possible outcome. This factor has also been analyzed in previous research (Glöckner et al., 2016; Wulff et al., 2017), and it made It possible to assess whether the difference between an outcome with a probability of .99 and a certain outcome is just quantitative (i.e., a difference of .01 in probability) or qualitative (i.e., a certain degree of risk versus certainty).

For the same reasons that were discussed on the previous analysis, data from the third experiment from Glöckner et al. (2016) was not included on this analysis.

e. Description task paradigm

In description-based tasks, two options are shown to the participants using text, graphs, or both. They have all the information available from the beginning: the outcomes values and probabilities are displayed for both options.

There are different ways of presenting the options on description-based tasks. The most common method is using probabilistic statements, for example: “Option A: win 10$ with p = .20, 5$ otherwise. Option B: win 7$ for sure.” Other methods of presenting description-based tasks involve the use of frequencies (e.g., “Option A: there are 10 tickets in this options, 2 that give you 10$ and 8 that give you 5$. Option B: there are 10 tickets in this options, all of them give you 7$.”), or graphics such as pie charts, tree graphs or open sampling (see Figures 3, 4 and 5).

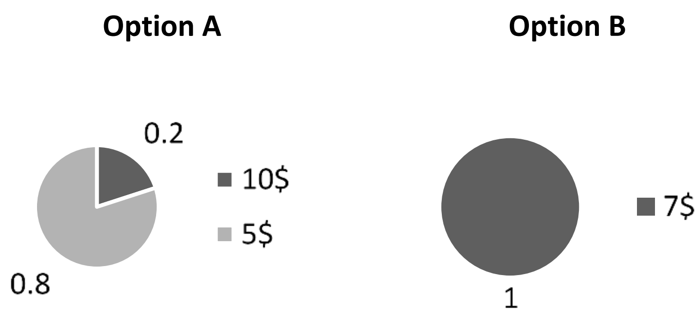

Figure 3. Options of a description-based task displayed as pie charts.

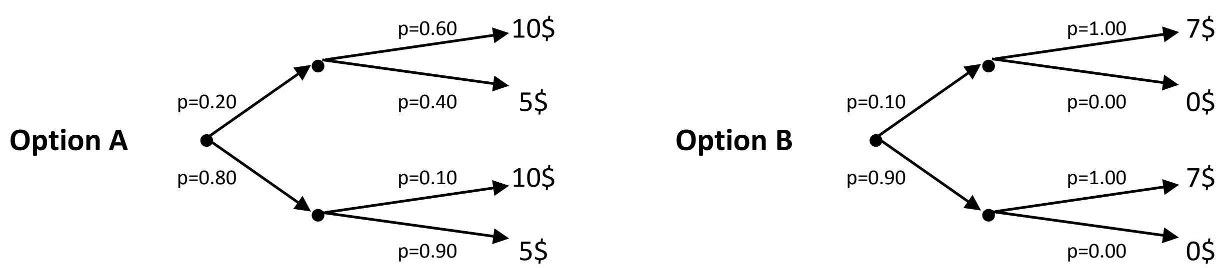

Figure 4. Options of a description-based task displayed as tree graphs.

Option A

5$	5$	10$	5$	5$
5$	10$	5$	5$	5$
5$	5$	5$	10$	5$
5$	5$	10$	5$	5$
10$	5$	5$	5$	5$

Option B

7$	7$	7$	7$	7$
7$	7$	7$	7$	7$
7$	7$	7$	7$	7$
7$	7$	7$	7$	7$
7$	7$	7$	7$	7$

Figure 5. Options of a description-based task displayed in an open sampling format.

A binary factor was used to test the magnitude of the description-experience gap depending on whether the options on the description-based task were presented using probabilities or other methods. Different authors have asserted that the way information is presented on description-based tasks may be an important factor on explaining the description-experience gap, as probability statements may be harder to understand than other methods (Gottlieb et al., 2007; Harman & Gonzalez, 2015; Hilbig & Glöckner, 2011).

-

f. Experience task paradigm

In experience-based tasks, two options are shown to the participants in a symbolic way (e.g., doors, bags, buttons, etc.). They know nothing about the outcomes values and probabilities. Each time a participant makes a choice, either just the consequence of the chosen option (partial feedback paradigm) or also the forgone consequence of the non-chosen option (full feedback paradigm) is presented. The outcomes probabilities can only be inferred by the participants, depending on the relative frequencies of the outcomes that they experience.

The sequence of the outcomes can be either controlled or truly random. Controlled sequences ensure that the participants take a sample of events that matches the *a priori* outcomes probabilities, while truly random sequences do not. For example, an option A that gives 10$ with a probability of .20 and 5$ otherwise, when sampled 20 times in a controlled sequence gives 4 times 10$ and 16 times 5$, while in a truly random sequence there are several possibilities.

There are two ways of presenting experience-based tasks. In the sampling format, the participants can sample outcomes, either for a limited or unlimited amount of times, from both options without consequences for them. Once this sampling phase ends, they have to “play for real”, and this last choice is considered the actual choice of the participants. On the other hand, the feedback format makes the participants choose a certain amount of times between two options, with real consequences since the very first choice; every choice is considered an actual choice of the participants.

A binary factor was used to test the magnitude of the description-experience gap depending on whether the experience-based tasks used a sampling or feedback method. It has been suggested that getting real consequences in every choice on experience-based tasks may be an important factor on explaining the description-experience gap, as the sampling paradigm may not be sufficient to produce underweighting of rare events (Camilleri & Newell, 2011b).

Missing data

If data from a particular publication was not available directly through electronic search, the authors were contacted in order to ask for it. Every author we contacted kindly put at our disposal their data.

Data synthesis

The meta-analysis was conducted by weighting the results of each study by their standard errors (Haidich, 2010). The standard deviation of each description and experience group was calculated in order to compute the standard error of standardized mean differences.

Results

Systematic review

a. Search strategy

From a total of 168 publications found using the electronic search, excluding repetitions, 28 were selected by both researchers, while 15 were only selected by one of them. Therefore, a moderate interrater reliability kappa statistic (McHugh, 2012) of .73 was obtained on the first selection phase of the publications, and all of the 43 publications were examined to assess their suitability with respect to our criteria.

From those publications, 19 met our criteria. Experimental data from Study 2 of the present Ph.D. thesis and 11 extra publications were added via references and additional search using Google Scholar and the database of Judgment and Decision Making and Behavioral Decision Making journals, adding up to a total of 31 papers to be analyzed (see Figure 6).

b. Summary of results

Table 1 shows the summary of the findings of each paper included on this review.

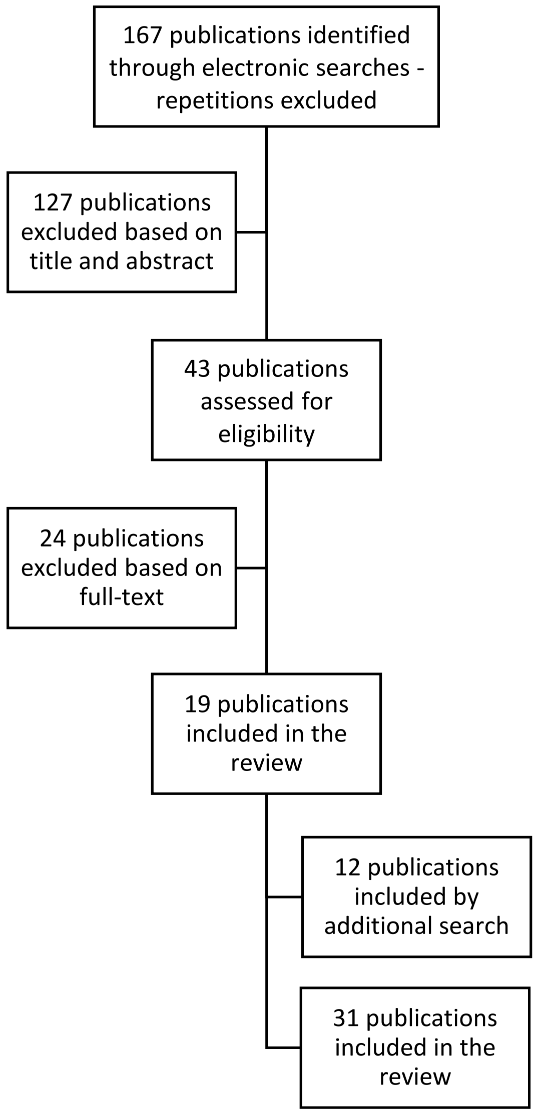

Figure 6. Publications inclusion process flow chart.

Table 1

Summary of the publications included on the meta-analysis and main characteristics.

Publication	N	Design	Description	Experience	Problems	$	Factors
Barron & Erev (2003)	281	Betw	Prob (1) Prob (play 100)	FB (last 100 of previous studies)[a]	5	$ / No	Memory biases
Hertwig et al. (2004)	100	Betw	Prob (1)	Samp (∞)	6	$	Memory biases Sampling bias
Weber et al. (2004)	165	With	Pie-chart with prob (1)	Samp (∞)	5	$	Coefficient of variation
Yechiam, Barron & Erev (2005)	78	Betw	Prob (1) Prob (play 100)	FB (100)	1	$	Experience with rare events
Gottlieb et al. (2007)	128	With	Prob (1) Freq (1) Open samp (1)	Samp contr (20 each)	16	$	Presentation format
Hau, Pleskac, Kiefer & Hertwig (2008)	233	Betw	Prob (1)	Samp (∞) Samp (∞) Samp (100)	6	$	Sampling bias Different processes
Rakow, Demes & Newell (2008)	240	Betw - With	Prob (1) Freq yoked Prob yoked	Samp (∞) Samp yoked passive Samp reverse yoked passive	12	$	Sampling bias
Camilleri & Newell (2009a)	80	Betw	Prob (1)	Samp control (∞)	8	Cr $	Different processes
Camilleri & Newell (2009b)	40	With	Prob (1)	Samp (∞)	10	Cr $	Sampling bias
Ungemach, Chater & Stewart (2009)	272	Betw	Prob (1)	Samp (∞) Samp (40 each) Samp contr (40 each)	6	Base / No	No memory biases No sampling bias
Erev et al. (2010)	320	Betw	Prob (1) Prob (1)	Samp (∞) FB (100) Samp (∞) FB (100)	120	$	Distance between cumulative payoff functions Sampling bias
Hau et al. (2010)	160	Betw	Prob (1) Prob yoked	Samp (50) Samp records (50)	12	Cr $	Rare events
Lejarraga (2010)	118	With	Tree gr (1)	Samp (∞)	7	$	Frequency judgments
Rakow & Rahim (2010)	324	Betw - With	Freq (1) Freq (1) Freq (1)	Samp (10 each) Samp (10 each) Samp (10 each)	16	No	Rare events
Camilleri & Newell (2011a)	138	Betw	Prob (1)	Samp (∞) Samp contr (∞) Samp contr (∞)	10	Cr $	Sampling bias

Camilleri & Newell (2011b)	120	Betw	Prob (1)	Samp (100) FB partial (100) FB (100)	4	Cr $	Repeated, consequential choices in the feedback paradigm
Hilbig & Glöckner (2011)	130	Betw	Prob (1) Open samp (1)	Samp (∞)	12	$	Presentation format
Lejarraga & Gonzalez (2011)	91	Betw	Prob (play 100) Prob complex (play 100)	FB (100)	2	$	Higher reliance on experienced outcomes
Ludvig & Spetch (2011)	83	With	Pie-chart (32)	FB partial contr (48) FB partial contr (48)	2	Cr	Different processes
Artinger, Fleischhut, Levati & Stevens (2012)	128	Betw	Prob (1)	Samp (25)	8	$	Different processes
Glöckner et al. (2012)	44	Betw	Prob (1)	Samp (∞)	37	$	Different processes
Kudryavtsev & Pavlodsky (2012)	75	With	Pie-chart (1)	FB (100)	10	$	Linear weighting of gains and losses
Camilleri & Newell (2013)	203	With	Prob (1) Prob (play 100) Prob (distr 100)	FB (41) FB (40 - play 100) FB (40 - distr 100)	32	Base	Sampling bias Loss aversion
Harman & Gonzalez (2015)	199	Betw	Prob (1) Freq (1)	FB (100)	2	Base	Properties of the options
Oeberst, Haberstroh & Gnambs (2015)	185	Betw	Prob (1) Prob (1)	Samp contr (50) Samp contr man (50) Samp contr (50) Samp contr man (50)	1	$ / No	Presentation format
Glöckner, Hilbig, Henninger & Fiedler (2016)	228	Betw	Prob (1) Prob (1) Prob (1)	Samp (∞) Samp (∞) Samp (∞)	113	$	Sampling bias Information asymmetry Regression to the mean
Kellen, Pachur & Hertwig (2016)	104	With	Prob (1)	Samp (∞)	114	Base $	Different processes
Ashby (2017)	324	With	Prob (1) Prob (1)	Samp (max 100) Samp (100)	21	$	Numeracy skills
Madan, Ludvig & Spetch (2017)	238	With	Pie-chart (32)	FB partial contr (48)	2	Cr $	Memory biases
Yoon, Vo & Venkatraman (2017)	116	With	Prob (1) Prob (1)	Samp (∞) Samp (∞)	14	Cr / Base	Different processes
Study 2	64	Betw	Prob (play 100) Comb (play 100)	FB (100)	2	Cr / Vou	Not comprehension of the descriptions

Note. Design: Betw = between-subjects; With = within-subjects. Description refers to the paradigm used in description-based tasks: Prob = text using probabilities; Prob complex = text using probabilities expressed in a complex way; Freq = text using frequencies; Comb = combination of text using probabilities, frequencies and expected value; Pie-chart = pie-chart graph; Tree gr = tree graph; Open samp = open sampling. Experience refers to the paradigm used in experience-based tasks: Samp = sampling; FB = full feedback; FB partial = partial feedback; contr = controlled experienced probabilities; man = manual task (as opposed to computer task). Both description and experience groups may have one or more of the following: Yoked = yoked with a participant from experience; Passive = computer makes the choices for the participant who stays watching; Reverse = reverse order; (X) = number of choices to be made; (play X) = the computer plays X time he chosen option for the participant; (distr X) = distribution between both options for X times. $ refers to the real gain for the participants: $ money contingent to behavior; Base = fixed amount of money; Cr = course credit; Vou = voucher depending on behavior; No = nothing. A publication can have more than one type of real gain. Two types of real gains separated by slash "/" means different types of gains for different groups.

[a] They compared the result of their description groups with the last 100 trials of their respective experience groups, in order to have the same number of trials (i.e. the studies they have on that same publication with experience groups had more than 100 trials).

A total of 31 publications conformed the systematic review. Most of them used undergraduate students as participants, adding up to a total of 5009 subjects. They were recruited from the Technion institute in Israel, the Ohio State University, the Rutgers University in New Jersey, the University of Basel in Switzerland, the University of Essex and the University of Warwick in England, the University of New South Wales in Australia, the University Pompeu Fabra in Spain, the Max Planck Institute for Research on Collective Goods, the University of Mannheim, the Berlin universities, the University of Bonn and Jena in Germany, the Carnegie Mellon University in Pennsylvania, the Temple University in

Philadelphia, the University of Alberta in Canada and the University of Minho in Portugal. Also, other studies included participants such as children, adolescents and adult general population from England, and American workers.

Most experimental designs are between-subjects, with 18 of 31 publications using it. Regarding the description and experience paradigms, the most used were the probability and the sampling paradigms, with 38 of 52 and 41 of 57 experimental groups, respectively. The median total number of tasks that these publications included was 8. Furthermore, 25 of 31 publications used money, alone or combined with other outcome, as the reinforcer of the experiment. Regarding the explaining factors to account for the description-experience gap, the variance is much bigger, as the most used factor by the authors is sampling with 8 mentions in the 31 publications.

Meta-analysis

Although forest plots are a common tool in meta-analyses, these graphs would not be functional on the present work because of the huge amount of data. Given that we worked not with study-level data but with individual problems, our forest plots would end having hundreds of lines. Therefore, a modified version of those was constructed in which a horizontal line for each study is drawn in a way to cover the same distance as the usual forest plot (i.e., from average - SE to average + SE). The main difference of this graph with standard forest plots is that individual study names and values are not shown, and that the diamonds representing the subgroup value are not enlarged depending on their standard error. The constant size across diamonds is due to the extremely low standard errors that were obtained, given that we worked with Cohen's *d* values, that would make most diamonds look more like vertical lines and difficult to see.

As the interpretation of I^2 commonly leads to mistakes (Borenstein, Higgins, Rothstein, & Hedges, 2017), we deem that the confidence interval of the difference remains as a better indicator of the actual effect size, as it is measured on the same scale than the individual studies, once converted to Cohen's *d*. Still, this indicator is provided together with the common effect size measure R^2 on each subgroup analysis.

a. Over- or under-weighting of the rare events

The results of the subgroup meta-analysis for the feedback paradigm data are shown in Table 2 and Figure 7. Desirable and undesirable subgroups comprised 126 and 191 results, respectively. Similar results were found at the aggregate data level comprising both sampling and feedback paradigms, and at the research from Wulff et al. (2017).

Let us remind that the description-experience gaps for this subgroup analysis were calculated by subtracting the proportion of choices containing the rare event on the description group from the experience group. Therefore, positive description-experience gaps are predicted when the rare event is undesirable (i.e., the undesirable rare event will be as-if overweighted in the description groups while underweighted in the experience groups), and vice versa for desirable rare events.

Table 2

Description-experience gap as a function of rare event value

Subgroup	Cohen's *d*	Lower	Upper	SE	I^2	R^2	Q-test	*p*
Desirable	-.558	-.651	-.466	.047				
Undesirable	.595	.503	.687	.047				
Overall	.022	-.043	.087	.033	56.15	.50		
Difference	1.153	1.023	1.283	.067			297.93	< .001

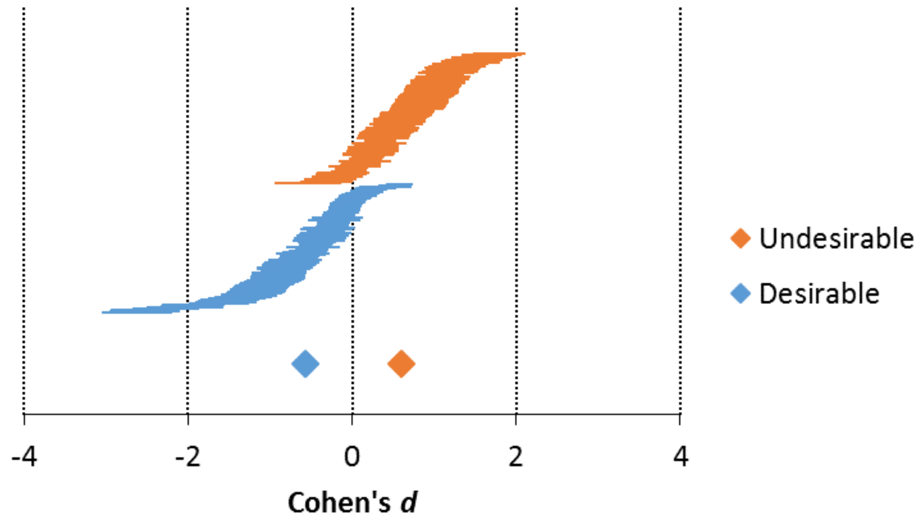

Figure 7. Description-experience gap as a function of rare event value. Horizontal lines represent the standardized mean differences and standard error of individual comparisons, sorted lower to higher for each subgroup. Diamonds and error bars (when visible) represent estimates and 95% confidence interval for each subgroup.

b. Domain

The results of the subgroup meta-analysis for the feedback paradigm data are shown in Table 3 and Figure 8. Gain, mixed and loss subgroups comprised 115, 50 and 99 results, respectively. The *Difference* row refers to the comparison between the gain and the loss domains. However, the direction of the difference was reversed at the aggregate data level comprising both sampling and feedback paradigms, and at the research from Wulff et al. (2017): while in the feedback paradigm gain domain gambles resulted in higher description-experience gaps, opposite results are found when both sampling and feedback, or only sampling, domains are considered.

Table 3

Absolute description-experience gap as a function of task domain

Subgroup	Cohen's *d*	Lower	Upper	SE	I^2	R^2	Z-test	*p*
Gain	.653	.572	.734	.041				
Mixed	.692	.563	.822	.066				
Loss	.520	.433	.607	.045				
Overall	.609	.555	.663	.028	13.62	.02		
Difference	.133	-.113	.192	.061			2.18	.015

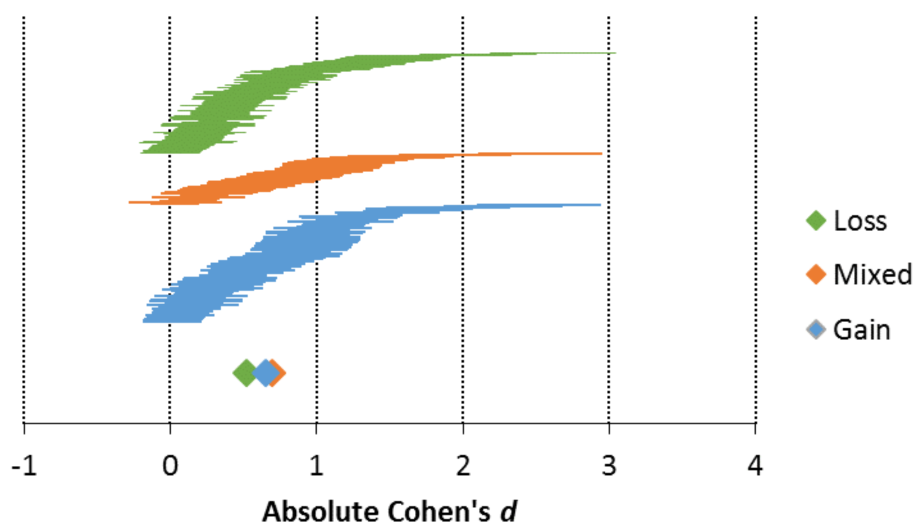

Figure 8. Absolute description-experience gap as a function of task domain. Horizontal lines represent the standardized mean differences and standard error of individual comparisons, sorted lower to higher for each subgroup. Diamonds and error bars (when visible) represent estimates and 95% confidence interval for each subgroup.

c. Probability of the rare event

The results of the meta-regression for the feedback paradigm data are shown in Table 4 and Figure 9. Similar results were found at the aggregate data level comprising both sampling and feedback paradigms. Moreover, the fit measures of the model are $Tau^2 = .139$, $I^2 = 70.30$.

Table 4

Absolute description-experience gap as a function of the rare event probability

Covariate	Coefficient	Lower	Upper	SE	z-test	*p*
Intercept	.791	.690	.892	.051	15.37	< .001
p(rare)	-1.418	-2.110	-.728	.352	-4.03	< .001

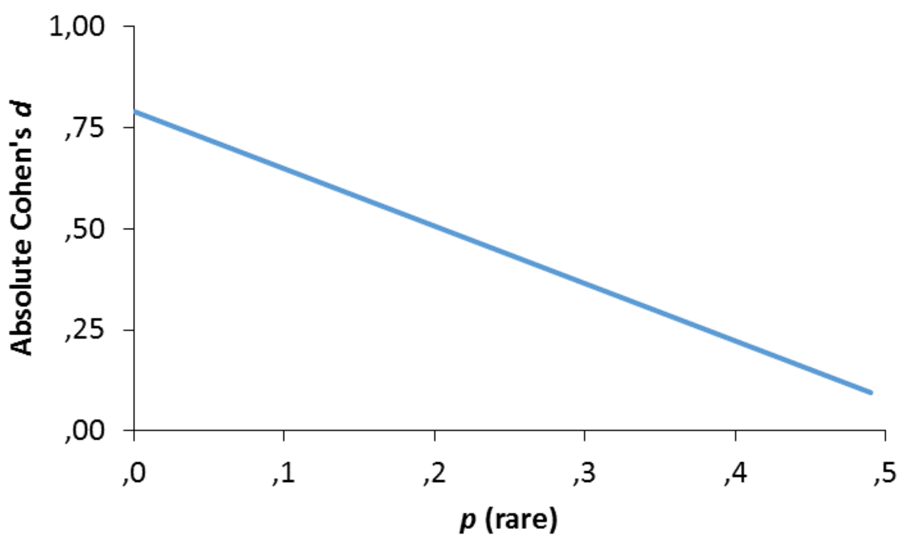

Figure 9. Absolute description-experience gap as a function of the rare event probability. Individual comparisons are not shown on the graph as they make it incomprehensible due to the huge amount of them.

d. Existence of a certain option

The results of the subgroup meta-analysis for the feedback paradigm data are shown in Table 5 and Figure 10. Safe and risky subgroups comprised 246 and 18 results, respectively. Similar results were found at the aggregate data level comprising both sampling and feedback paradigms, and at the research from Wulff et al. (2017). However, at the aggregate data level, a higher effect size was obtained, R^2 = .22. This difference on the effect size was probably due to the small amount of individual comparisons available on the feedback paradigm.

Table 5

Absolute description-experience gap as a function of the existence of a certain option

Subgroup	Cohen's *d*	Lower	Upper	SE	I^2	R^2	Q-test	*p*
Safe	.629	.573	.685	.029				
Risky	.386	.197	.575	.097				
Overall	.609	.555	.663	.028	13.50	.02		
Difference	.243	.046	.441	.101			5.79	.016

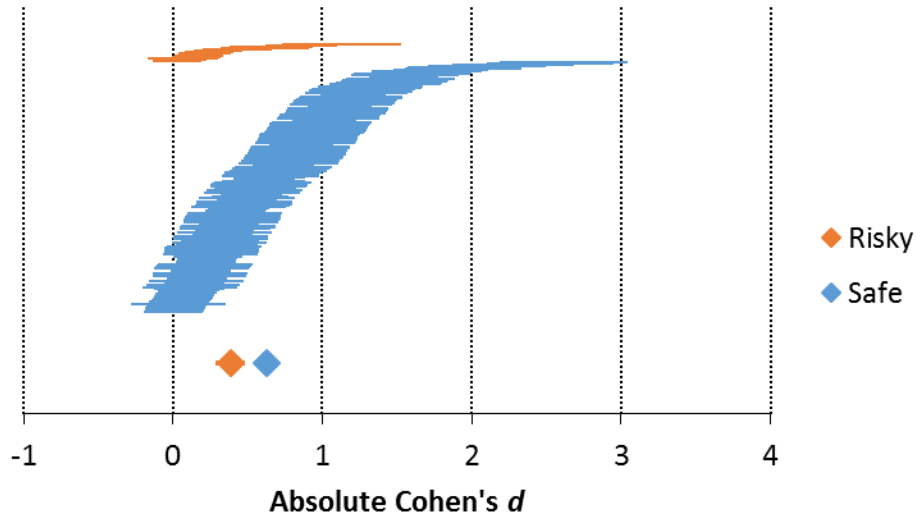

Figure 10. Absolute description-experience gap as a function of the existence of a certain option. Horizontal lines represent the standardized mean differences and standard error of individual comparisons, sorted lower to higher for each subgroup. Diamonds and error bars (when visible) represent estimates and 95% confidence interval for each subgroup.

e. Description task paradigm

The results of the subgroup meta-analysis for the aggregate data comprising both sampling and feedback paradigms are shown in Table 6 and Figure 11. Probabilities and other subgroups comprised 244 and 20 results, respectively. Similar results were found at the aggregate data level comprising both sampling and feedback paradigms.

Table 6

Absolute description-experience gap as a function of the description task paradigm

Subgroup	Cohen's *d*	Lower	Upper	SE	I^2	R^2	Q-test	*p*
Probabilities	.638	.583	.693	.028				
Other	.304	.131	.477	.089				
Overall	.607	.554	.660	.027	17.01	.08		
Difference	.334	.152	.515	.093			12.86	< .001

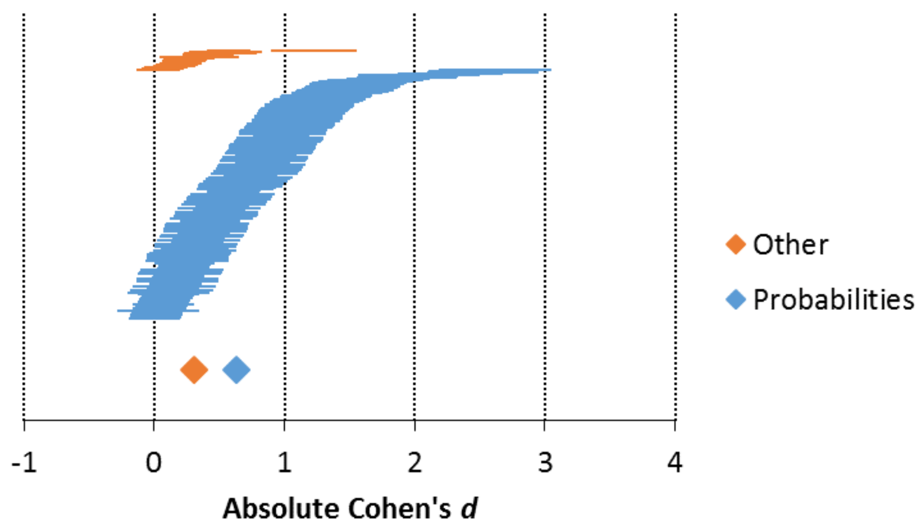

Figure 11. Absolute description-experience gap as a function of the description task paradigm. Horizontal lines represent the standardized mean differences and standard error of individual comparisons, sorted lower to higher for each subgroup. Diamonds and error bars (when visible) represent estimates and 95% confidence interval for each subgroup.

f. Experience task paradigm

The results of the subgroup meta-analysis for the aggregate data comprising both sampling and feedback paradigms are shown in Table 7 and Figure 12. Sampling and feedback subgroups comprised 704 and 264 results, respectively.

Table 7

Absolute description-experience gap as a function of the experience task paradigm

Subgroup	Cohen's *d*	Lower	Upper	SE	I^2	R^2	Q-test	*p*
Sampling	.348	.323	.373	.013				
Feedback	.584	.541	.627	.022				
Overall	.407	.385	.428	.011	21.39	.10		
Difference	.236	.186	.286	.026			85.21	< .001

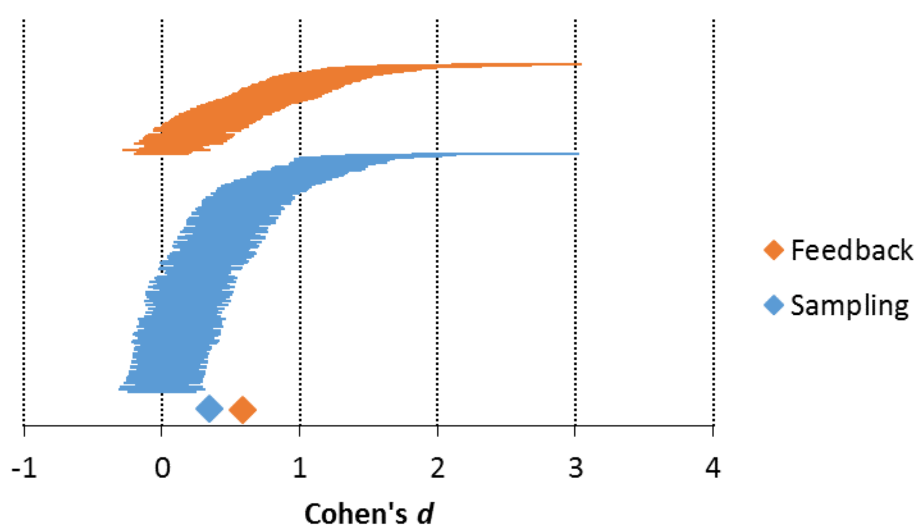

Figure 12. Absolute description-experience gap as a function of the experience task paradigm. Horizontal lines represent the standardized mean differences and standard error of individual comparisons, sorted lower to higher for each subgroup. Diamonds and error bars (when visible) represent estimates and 95% confidence interval for each subgroup.

Discussion

Our systematic review revealed the most common experimental characteristics of the studies on the description experience gap: between-subjects design, probability and sampling paradigms and using a total of 8 problems and money as reinforcer, while sampling biases is the most appealed explaining factor. In order to consider the quantitative findings on the description-experience gap, our meta-analysis will be discussed on the following paragraphs.

The meta-analysis focused on studying the main effects of different factors on the description-experience gap: over- and under-weighting of rare events in description- and experience-based tasks, domain of the task, probability of the rare event, existence of a certain option, description task paradigm and experience task paradigm. All of them revealed to be statistically significant factors on the description-experience gap using the feedback paradigm and at the aggregate data level comprising both sampling and feedback paradigms. These results and previous hypothesis are contrasted below.

General discussion

Hertwig, Barron, Weber and Erev (2004)'s claim that rare events function as-if they are overweighted in decisions from description while underweighted in decisions from experience finds support in this meta-analysis: when facing a choice, options containing positive rare events are more preferred in decisions from description than in decisions from experience, and vice versa for options containing negative rare events. This result is coherent with previous reviews (Erev & Roth, 2014; Hau et al., 2010; Hertwig, 2012; Rakow & Newell, 2010) and the other existing meta-analysis on the field (Wulff et al., 2017) and remains as the fundamental factor when describing the description-experience gap. In Liberman and Tversky (1993) terms, the pattern encountered in experience-based tasks may be called over-extremity, as the probabilities are "pushed" towards the extremes (i.e., 0 and 1), while under-extremity implies the pattern encountered in description-based tasks:

probabilities are “pushed” near to .5. Different theories have been constructed in order to understand the psychological foundations of this functions. Some of the most popular models are: optimistic overconfidence, confirmatory bias, case-based judgments, ecological models and error models (Griffin & Brenner, 2008).

Loss domain was shown to cause a lower description-experience gap compared to gain domain on the feedback paradigm. Although the reverse results were found when the full range of data was taken into account, and when using the sampling paradigm (Wulff et al., 2017). Further interesting research could evaluate this effect so it could be disentangled from studies sampling issues (i.e., there are too few studies evaluating problems on the loss domain using the feedback paradigm). The results obtained at the aggregate data level are coherent with Cumulative Prospect Theory prediction that departure from linearity is more pronounced for negative, compared to positive, outcomes (Tversky & Kahneman, 1992). As described on their theory, people tend to be risk averse for gains while risk seeking for losses: they will choose more often the certain option when making decisions in the gain domain. This pattern of choices will lead to a lesser variability between groups, as there is no rare event to be over- or under-weighted in certain options, which will lead itself to a reduction in the description-experience gap.

When we compared the absolute description-experience gap as a function of the rare event probability, we found linear model to be statistically significant and similar to what was found on the sampling paradigm (Wulff et al., 2017). This result confirms that, the rarer the rare event, the bigger the description-experience gap. This result and the previously explained overweighting of rare events support the Cumulative Prospect Theory, as it states that “departures from linearity, which violate expected utility theory, are most pronounced near the edges” (Tversky & Kahneman, 1992). Some authors have stated that the presence of rare events is what causes the description-experience gap (Hau et al., 2010; Rakow & Rahim, 2010). Therefore, the lower probabilities of the rarest event on the gambles should be inversely related to the description-experience gap, as it happened to be in our meta-analysis.

Our meta-analysis showed another interesting result: the description-experience gap is bigger when the problem contains a certain option. This result has been studied in the sampling paradigm and yielded similar results (Wulff et al., 2017). Researchers working in mathematical models of the description-experience gap may find it useful, as this means that there is a qualitative difference between an option that gives a particular amount of points/money with a probability of .99 and the same option involving a certain event (p = 1.00), therefore giving its origin to two different functions. An example of this qualitative difference is seen when people buy lottery tickets: they are willing to expend money in a gamble with such a low favorable probability, but of course they would not play it if the probability was absolute zero. The concepts of expected value and expected utility are of great importance here, as the only way of making attractive such lotteries, giving their extremely low favorable probabilities, is by giving huge prizes (D. Bernoulli, 1738/1954; Pascal, 1669/2002). Furthermore, when both options comprise more than one possible outcome, they would be applied with the determinants of the description-experience gap and thus they may cancel out (Wulff et al., 2017).

Two factors were exclusively studied on our meta-analysis regarding description and experience tasks paradigms. They will be discussed on the following paragraphs.

When both types of description task paradigms were compared, results showed that tasks described using methods other than probabilities yielded a lower description-experience gap than those that used probabilities. A plausible explanation of this difference may be that participants do not generally understand well probabilistic statements. Being that the case, when other methods are used, participants get a more objective comprehension of the probabilities and their choice behavior is affected, as has been shown in previous research (Gottlieb et al., 2007; Harman & Gonzalez, 2015; Hilbig & Glöckner, 2011).

Data comparison between experience task paradigms using sampling versus feedback revealed a reduction on the description-experience gap when the task involved sampling. This result is coherent with results obtained by Camilleri and Newell (2011b) showing that using the feedback paradigm is the best method to observe underweighting of rare events. Furthermore, it is fundamental to remark the importance of an explanation comprising the behavior of participants in experience-based tasks to explain the description-experience gap, as the explaining factor which remains the most recurrent on the literature is sampling behavior (Camilleri & Newell, 2011a, 2013; Erev et al., 2010; Glöckner et al., 2016; Hau et al., 2008; Hertwig et al., 2004; Rakow et al., 2008). The differential impact of both types of consequences –real vs. virtual– should be the starting point of such an explanation.

Limitations

Main effects and their interaction with experience paradigm were analyzed on the present meta-analysis. The interactions were conducted by comparing results obtained on the feedback paradigm with those at the aggregate data level comprising both paradigms, and the meta-analysis focusing on the sampling paradigm (Wulff et al., 2017). It still remains as an interesting idea for future work to evaluate further possible interactions between other factors. Also, these extra interactions could probably help the analyses yield higher effect sizes.

The greatly diverse list of places where the studies have been conducted sure is adding extra noise and variability to the meta-analysis. Higher effect sizes could be found if population characteristics were treated as additional factors. Nonetheless, this diversity is also adding validity to the meta-analysis, as it proves that the description-experience gap is indeed a robust effect that appears in a great variety of subjects around the world.

Conclusion

Some fundamentals of the description-experience gap have been analyzed and corroborated on the present work, namely: the as-if weighting functions of rare events, the task domain, the probability of the rare event, the presence of a certain option, and the paradigms in both the description- and experience-based tasks.

In order to conclude at a more theoretical level, let us suggest a plausible property of the description-experience gap that has not been explicitly arisen yet to our knowledge. Two different results will be used to explicitly support this hypothesis: different factors appealed in the literature as the causes of the description-experience gap, and the modulation of the effect by autonomous sampling.

When looking for explanations of the description-experience gap, the reference model that can be found (Tversky & Kahneman, 1992) is a descriptive one. Furthermore, the most appealed explaining factor is sampling, together with other authors' explanations referring the presentation format or the presence of rare events –which are all part of the methodology of the task itself–. This is, sampling bias has been asserted to be a fundamental factor underlying the description-experience gap (Camilleri & Newell, 2009b, 2011a, 2013; Erev et al., 2010; Glöckner et al., 2016; Hau et al., 2008; Hertwig et al., 2004; Rakow et al., 2008). Furthermore, it has been demonstrated that it is modulated by autonomous sampling in tasks using the sampling paradigm (Wulff et al., 2017) (i.e., it is further reduced when sampling behavior is controlled by the researchers). This explanation of the description-experience gap relies on the properties of the task itself too.

Both arguments have in common that they are searching for the causes of the cap in the pure methodological properties of the tasks, instead of in other psychological functions. Our hypothesis about the description-experience gap rely on these assertions to suggest that the procedural differences between the description- and the experience-based tasks are the causes underlying the phenomenon (i.e., a phenomenon that does not rely on other psychological mechanisms, as different "processes" for description and experience, or "memory biases"). This is, the linguistic, and therefore symbolic, properties of the description groups, contrasted with the contingencies generated on the experience groups may be the ultimate factors causing the discrepancies between them. This assertion has been widely studied on the research field of instructional control and similar conclusions were obtained (Catania et al., 1982; Galizio, 1979; Matthews et al., 1985, 1977; Shimoff et al., 1981). However, there are empirical contradictions between the results obtained on the mentioned studies on instructional control and two studies on the description-experience gap (Jessup et al., 2008; Lejarraga & Gonzalez, 2011). The following experiments will serve to disentangle this contradiction.

Study 2

Type of description[2]

The present study aims to reveal contradictory results obtained on two different fields; particularly from two studies conducted on the description-experience gap field showing that descriptions are neglected when personal experience is available (Jessup et al., 2008; Lejarraga & Gonzalez, 2011), and several others conducted on the instructional control field getting to the opposite conclusion (e.g., Catania et al., 1982; Galizio, 1979; Hackenberg & Joker, 1994; Matthews et al., 1985, 1977; Shimoff et al., 1981). To account for this contradiction, we hypothesized that participants from the former studies relied on their experience rather than on the descriptions because of the difficult, demanding nature of the probabilistic descriptions they faced. Enriched descriptions were created in our experiment to assess the contribution of this factor to the differential influence of the descriptions in choice behavior. Nonetheless, our hypothesis did not find support in the results and further research is needed to account for the aforementioned contradiction.

[2]This chapter reproduces the publication: Viudez, A., Keating, J., & Arantes, J. (2018). The description-experience gap and its relation to instructional control: Do people rely more on their experience than in objective descriptions? *Journal of Negative and No Positive Results, 2*(12), 667–675. Minor modifications were made to avoid overlap between different chapters.

Introduction

Between the different Psychology paradigms, not only theoretical but also empirical differences are found: depending on the definition of the science itself and its study field, the experimental interests will vary greatly among all of them. However, there are certain overlaps on the research fields that could be of benefit to all of us if revealed so we could take advantage of what other researchers, even with different theoretical coordinates, have already done.

The present work aims to reveal contradictory results obtained on two different fields; particularly from two studies conducted on the description-experience gap field showing that descriptions are neglected when direct experience is available (Jessup et al., 2008; Lejarraga & Gonzalez, 2011), and several others conducted on the instructional control field getting to the opposite conclusion (e.g., Catania et al., 1982; Galizio, 1979; Hackenberg & Joker, 1994; Matthews et al., 1985, 1977; Shimoff et al., 1981). In order to disentangle this contradiction, we studied how the descriptions of the options are presented in the description-experience gap experiments to understand why they do not affect participants' behavior.

Little is known about the description-experience gap paradigm when both descriptions and direct experience are available. On that sense, Jessup, Bishara, and Busemeyer (2008) divided the participants in two groups and found that the presence of feedback on repeated decisions from description (*mixed group*) altered the choice behavior compared to the group without feedback *(description group*). Subjects on the *mixed group* were shown the possible outcomes and probabilities, at the same time that were allowed to play the same gamble repeatedly so they could learn the outcomes distributions from their experience too. Subsequently, Lejarraga and Gonzalez (2011) conducted an experiment using three groups: *description, experience* and *mixed*. The results showed that choice behavior of the subjects from the *mixed group* was statistically different from those of the *description group* and almost identical to those of the *experience group*. The authors concluded that when both sources of information are available, individuals neglect the descriptions (Lejarraga & Gonzalez, 2011).

Nonetheless, these results regarding a higher control on behavior of schedule contingencies compared to descriptions (Jessup et al., 2008; Lejarraga & Gonzalez, 2011) seem to contradict aforementioned research on instructional control and research in standard choice tasks (Barron, Leider, & Stack, 2008; Fantino & Esfandiari, 2002) or the prisoner dilemma (Baker & Rachlin, 2001). This inconsistency serves as the theoretical context for the present study, as an explanation to account for it is needed.

To account for this contradiction, we hypothesized that participants from the *mixed groups* from Jessup et al. (2008) and Lejarraga and Gonzalez (2011) relied on their experience rather than on the descriptions because of the difficult, demanding nature of the standard probabilistic descriptions of the options. Indeed, Lejarraga and Gonzalez (2011) also had groups with descriptions that were very complex and difficult to understand, and found no difference between these groups and the original ones (Lejarraga & Gonzalez ,2011). This result supports our hypothesis, as subjects treat the standard descriptions of the problems the same way they do with descriptions that are clearly hard to understand in order to make their choices.

Additionally, some authors have demonstrated that the presentation format of the description-based tasks affects choice behavior on the description-experience gap (Gottlieb et al., 2007; Harman & Gonzalez, 2015; Hilbig & Glöckner, 2011). Therefore, we hypothesize that if participants on the *mixed group* had a better comprehension of the descriptions, their behavior would be more similar to the *description* rather than the *experience group,* and therefore being coherent with previous research on instructional control.

Method

Two choice tasks involving probabilities were presented to the participants in a 3 x 2 design with 3 levels of source of information available –only description, only experience or both– and 2 levels of information available on the descriptions –original or enriched–. The *experience group* is insensitive to the second factor, as there are no descriptions given on it. Therefore, five groups were formed.

The *description*, *experience* and *mixed groups* were used to replicate the original results from Lejarraga and Gonzalez (2011) in order to not confound population effects, while the remaining two groups –*enriched description* and *enriched mixed*– served to test the effect of enriched descriptions on the subjects' choice behavior.

In tasks involving typical, probabilistic descriptions, some participants may be able to calculate the relative frequencies of each outcome and the expected value of the options and make their choice based on this information, while others cannot. By using enriched descriptions including this information we aimed to turn this extraneous variable into a constant that we deemed fundamental.

Participants

Our sample included 104 students from the University of Minho (85% females, 86% Psychology students). Participants earned a course credit and entered on a raffle of a 20€ FNAC voucher for their participation in the study.

Materials

Participants faced two choice tasks used by Jessup et al (2008), and Lejarraga and Gonzalez (2011) on a computer screen, presented in random order. Nonetheless, due to a design failure, 72% of the subjects (*n* = 75) faced problem A before problem B.

In each task, the subjects faced a *safe* and a *risky option* associated to two buttons whose position –left or right– was randomized for each participant. The *safe option* would give 3 points for sure in both problems, while the *risky option* would give 4 points with a probability of 80% in problem A and 64 points with a probability of 5% in problem B. At the end of the experiment, participants had to fill a questionnaire about task comprehension and gambling habits. The experimental program was written using the OpenSesame software.

Procedure

Participants were randomly assigned to one of five groups: *description, experience, mixed, enriched description* and *enriched mixed*. On the task' instructions, participants were told that they would do two choice tasks to earn points that would allow them to win a €20.00 FNAC voucher.

Groups differed on the presentation format of the problems (see Table 8). We used the same methodology than Lejarraga and Gonzalez (2011) to make a fair comparison between publications results. Specifically, the *description* group was presented with the probabilistic description of each problem and had to choose one option. It was said to the participants that the computer would play that option 100 times. The *experience* group faced two unlabeled buttons, and had to choose between both options 100 times. The *mixed* group was presented with both the probabilistic description of the problems, and the experienced outcomes, as they had to choose between them 100 times. The remaining two groups –*enriched description* and *enriched mixed*– were similar to the *description* and the *mixed groups*, respectively, with one exception: the description of the problems were more detailed, as they involved probabilities, frequencies and expected values.

Every choice made by the participants had real consequences for them, as they got as many points as the sum of the outcomes of the 100 trials for both choice tasks. Participants saw, highlighted in red, the amount of points they got in each trial, and also the amount of points they missed on the other option.

Table 8

Description of the problems for different groups (original in Portuguese)

Problem A	**Problem B**
Experience group descriptions	
"A - B"	"A - B"
Description and *Mixed groups* descriptions	
Safe option: win 3 points for sure. Risky option: win 4 points with an 80% chance or win 0 points otherwise.	Safe option: win 3 points for sure. Risky option: win 64 points with a 5% chance or win 0 points otherwise.
Enriched description and *Enriched mixed groups* descriptions	
Safe option: win 3 points for sure. (You will get 3 points every time you select this option, so you will get a total of 30 points for every 10 times you choose this option.) Risky option: win 4 points with an 80% chance or win 0 points otherwise. (On average, you will get 4 points 8 out of 10 times you select this option and 0 points the remaining 2 times. So, on average, you will get a total of 32 points for every 10 times you choose this option.)	Safe option: win 3 points for sure. (You will get 3 points every time you select this option, so you will get a total of 30 points for every 10 times you choose this option.) Risky option: win 64 points with a 5% chance or win 0 points otherwise. (On average, you will get 64 points 1 out of 20 times you select this option and 0 points the remaining 19 times. So, on average, you will get a total of 64 points for every 20 times you choose this option.)

Results

Figure 13 shows average proportion of choices of the risky option in blocks of 25 trials for problem A (left panel) and problem B (right panel). Results from Lejarraga and Gonzalez (2011) are plotted on the first row to allow visual comparison, while our groups using the original and enriched descriptions are plotted on the second and third row, respectively.

Results from Lejarraga and Gonzalez (2011) were successfully replicated with the exception of the *description group:* our participants' proportion of choices of the risky option –*R-rate*– was lower on problem A while similar on problem B (see upper and middle rows of Figure 13 to compare Lejarraga and Gonzalez (2011) results with our replication).

On the other hand, results showed that introducing more information about the options on the descriptions did not have an effect on choice behavior for either the *description* or the *mixed group* for problem A (see left panel of middle and lower rows of Figure 13). Nonetheless, the type of description did have a differential effect for problem B, as we found a significant description-experience gap as revealed on a one way ANOVA when using the original descriptions, $F(2, 57) = 7.05$, $p = .002$; but no difference between conditions when using our enriched descriptions, $F(2, 61) = 1.14$, $p = .327$.

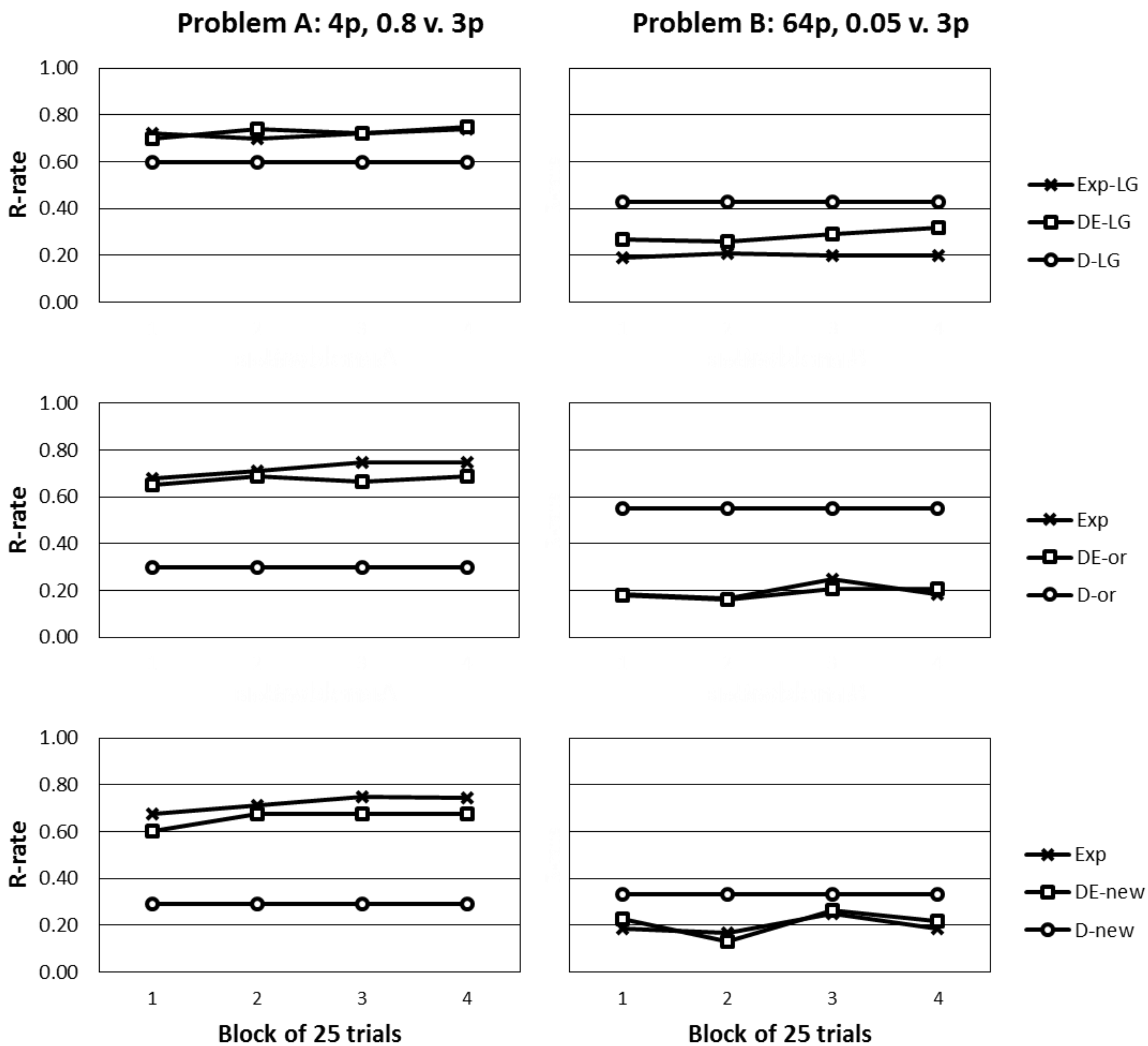

Figure 13. Choice behavior expressed as risky rate in blocks of 25 trials by problems and conditions. Left panels show data for problem A and right panels for problem B. Upper row shows data from the experiment of Lejarraga and Gonzalez (2011), middle row shows data for original descriptions and lower row shows data for the new, enriched descriptions.

More specifically, choice behavior was found to be similar for the *mixed* and *experience groups* in problem A both with the original descriptions, $F(1, 38) = 0.91$, $p = .346$; and with our enriched descriptions, $F(1, 38) = 1.51$, $p = .227$. On the other hand, the *mixed* and *description groups* were found to be different both with the original descriptions, $F(1, 38) = 10.67$, $p = .002$; and with our enriched descriptions, $F(1, 42) = 10.94$, $p = .002$.

In problem B, choice behavior was also found to be similar for the *mixed* and *experience groups* both with the original descriptions, $F(1, 38) = 0.02$, $p = .900$; and with our enriched descriptions, $F(1, 38) = 0.05$, $p = .829$. *Mixed* and *description groups* were found to be different when using the original descriptions, $F(1, 38) = 8.15$, $p = .007$; but equal when using our enriched descriptions, $F(1, 42) = 1.15$, $p = .289$.

Regarding the gambling habbits, 19% of our participants responded that they never gamble, while 31%, 33%, 11% and 6% of them responded they play 1, 2, 3 or 4 or more types of gambles, respectively. From those who gamble, the most common frequency was “ocassionally”, followed by “once a month” and “once a week”: 79%, 16% and 5%, respectively.

Discussion

Our study analyzed the effect of accurate, enriched descriptions on choice behavior. In two different problems, five conditions were compared: *description, experience, mixed, enriched description* and *enriched mixed*. Enriched descriptions affected *description groups* in problem B, but had no effect in problem A. On the other hand, they did not affect *mixed group* in problem A, while the assessment of their function in problem B could not be properly assessed, as the variance between groups was null (i.e., it was not possible to compare data from the *mixed group* with *description* and *experience* groups in a differential way, as no significant difference between them was found).

While results from our *mixed* and *experience groups* replicated those from Lejarraga and Gonzalez (2011), our *description group* showed a different choice behavior. Lower and higher preferences for the risky option were found in problem A and B, respectively, when comparing our groups with theirs. This difference, probably caused by a population effect, points out the importance of replicating previous results (Open Science Collaboration, 2015), as not doing it before introducing experimental modifications could induce biased comparisons.

The enriched descriptions were proved to be useful for our purpose: turn the information that the participants extract from the descriptions of the options into a constant. The fact that a difference was found between *description* and *enriched description groups* in problem B, shows that, with the original descriptions, the participants could not extract all the information by themselves. This result is coherent with previous studies showing the difficult nature of probabilistic statements to be understood by participants (Gottlieb et al., 2007; Harman & Gonzalez, 2015; Hilbig & Glöckner, 2011) and with another result from Lejarraga and Gonzalez (2011) showing that participants' choice behavior did not change when descriptions of the options were highly complex and therefore harder to be understood. Therefore, more research on the participants' understanding of the descriptions is needed to properly assess how descriptions should be presented in order to maximize their compressibility. Working with graphical descriptions of the options, some authors have suggested that the presentation format is the main explaining factor of the description-experience gap (Gottlieb et al., 2007; Hilbig & Glöckner, 2011).

Regarding our hypothesis about the contradiction between the results obtained on the description-experience gap (Jessup et al., 2008; Lejarraga & Gonzalez, 2011) and classic studies on instructional control (Baker & Rachlin, 2001; Barron et al., 2008; Catania et al., 1982; Fantino & Esfandiari, 2002; Galizio, 1979; Hackenberg & Joker, 1994; Matthews et al., 1985, 1977; Shimoff et al., 1981) being caused by incomprehensible descriptions of the tasks, it did not find support in the results. We expected to find the *enriched mixed group* to be similar to the *enriched description* while different from the *experience group* when using enriched descriptions. However, problem A clearly showed the same pattern of results on the *mixed groups* regardless of the type of descriptions that were used, while problem B actually did not showed enough variance to look for differences between *mixed group* and the others: as the *enriched description* and *experience groups* did not show statistically significant differences between them, there was not any possibility to test our hypothesis in this problem, because of the lack of variance between both groups. On this scenario, two possibilities arise: either the enriched descriptions would have had an effect in problem B but the small range on the results prevented it to appear, or they would not have had an effect anyway.

If the first scenario is the case, there would be a clear difference on the importance of the descriptions on a *mixed group* between problems A and B, suggesting that maybe just in some scenarios, the descriptions override the experience, and further research defining those scenarios would be needed in order to understand the factors behind it.

Limitations of our study include uneven proportions of subjects' gender and order of presentation of the problems, and the impossibility to assess whether our enriched descriptions made subjects' choice behavior from the *mixed group* in problem B more similar to either *description* or *experience group.* Using additional problems may shed light to this question in further research.

If indeed the comprehension level of the descriptions does not make a difference in any of the suggested scenarios, we would still continue with the same question that gave rise to this study: why do we find this contradiction between several studies demonstrating the bigger importance of descriptions compared to experience (Baker & Rachlin, 2001; Barron et al., 2008; Catania et al., 1982; Fantino & Esfandiari, 2002; Galizio, 1979; Hackenberg & Joker, 1994; Matthews et al., 1985, 1977; Shimoff et al., 1981) with those that show the opposite result (Jessup et al., 2008; Lejarraga & Gonzalez, 2011)? Our experiment was not able to disentangle this problem.

In conclusion, the present study showed that (1) the description-experience gap was found in both tasks; (2) choice behavior from our *mixed group* replicated the results from Lejarraga and Gonzalez (2011); (3) using enriched descriptions had a significant effect on the *description group* in one of the tasks; (4) further research is needed in order to understand participants' comprehension of the descriptions of the options; and (5) our hypothesis for the contradiction in the results from the description-experience gap with those from the instructional control field did not find support in our results. Studies 3 and 4 will evaluate how the type of schedule and the relative expected values between options affect this contradiction.

Studies 3 and 4

Type of schedule and relative expected values[3]

The aim of Studies 3 and 4 is to solve the contradiction in the results that also served as theoretical framework for Study 2. Particularly, we compared two studies conducted on the description-experience gap field showing that descriptions are neglected when personal experience is available (Jessup et al., 2008; Lejarraga & Gonzalez, 2011), with one study that used concurrent variable-interval schedules getting to the opposite conclusion (Takahashi & Shimakura, 1998). Two factors were studied: the type of schedule, and the relative expected values between options. The present work showed that (1) positive evidence of instructional control was found in a choice task with probability schedules and different expected values between options; (2) negative evidence of instructional control was found in a choice task with variable-interval schedules and similar expected values between options; and (3) when taking these results together with previous research, relative expected values was demonstrated to be a fundamental factor on understanding the presence of instructional control in choice tasks. We conclude that the relevance of this factor relies on its capacity to make participants' decisions easier by adding descriptions: participants better discriminate which is the best option.

[3]This chapter reproduces the document: Viudez, A., Keating, J., Arantes, J., & Martinez, H. (2017). Instructional control in choice tasks: the relation between reinforcement probability and time, and relative expected values. *Manuscript submitted for publication.* Minor modifications were made to avoid overlap between different chapters.

General introduction

The inconsistency on the results that served to contextualize Study 2 remains as the rational of Studies 3 and 4. Study 2 demonstrated that this contradiction is not caused by a poor comprehension of the descriptions. However, the experimental procedures of the cited experiments on the description-experience gap and instructional control are so different that it would be difficult to disentangle what is causing this discrepancy.

Nonetheless, one study from the instructional control field (Takahashi & Shimakura, 1998) used a more similar procedure to those of the description-experience gap, as it explicitly examined choice behavior. The authors used concurrent variable-interval schedules to assess choice behavior by evaluating its conformity to the matching law (Baum, 1974; Herrnstein, 1961) and found evidence of instructional control. Therefore, it was possible to assess the variables underlying it because of the methodological similarities.

Two main factors make the procedures of the description-experience gap tasks (Jessup et al., 2008; Lejarraga & Gonzalez, 2011) and the concurrent variable-interval schedules (Takahashi & Shimakura, 1998) differ: the type of schedule, and the relative magnitude of the reinforcers. More specifically:

1. While using concurrent variable-interval schedules makes the probability of reinforcement on each option depend on time, tasks from the description-experience gap field involve probability schedules that are independent of time (i.e., every time the participant chooses a particular prospect, the probability of getting reinforcement remains the same).
2. On the experiment from Takahashi and Shimakura (1998), both options delivered the same reinforcer, leading to different expected values due to the differences on reinforcers frequencies. However, on the description-experience gap experiments the magnitudes of the reinforcers are different so that expected values from both options remain similar.

Therefore, having only two procedural differences between experiments from both research fields, the evaluation of their contribution to instructional control is straightforward: a two-factor table with two conditions on each factor was constructed to disentangle the situation of these differences so far (see Table 9). There is negative evidence of instructional control when the task involves probability schedules using options with similar expected values –description-experience gap tasks– (Jessup et al., 2008; Lejarraga & Gonzalez, 2011). On the other hand, there is positive evidence of instructional control when the task involves variable-interval schedules using options with different expected values –concurrent variable-interval schedules– (Takahashi & Shimakura, 1998). Consequently, the remaining two possibilities shall serve as the experimental procedures for Studies 3 and 4 in order to disentangle the specific contribution of each factor to the instructional control in choice tasks.

Table 9

Preliminary evidence of instructional control as a function of expected values of the options and type of schedule

	Probability	**Variable-interval**
Similar expected values	Negative evidence	Experiment 2
Different expected values	Experiment 1	Positive evidence

Taking into account the information given in Table 9, Study 3 will evaluate instructional control on a task using probability schedules and options with different expected values (e.g., a choice task between a prospect that gives 1 point with 40% of chance and a prospect that gives 1 point with 20% of chance). The aim of Study 4 is to investigate instructional control on a task using variable-interval schedules and options with similar expected values (e.g., a concurrent VI 60-s VI 180-s schedule where the former gives a reinforcer of 20$ while the later gives a reinforcer of 60$). Results will be evaluated at a group level on Study 3 while at an individual level on Study 4, following adequate data analysis for each experimental methodology.

Once both experiments are finished we will be able to complete the schema in Table 9 in order to know which factor is causing the contraction in the results regarding instructional control.

Study 3

Method

Two choice tasks involving probability schedules were presented to the participants in a single-factor, between-subjects design with two levels of source of information available –only experience or experience and description–. Both options delivered reinforcers of similar magnitude, namely 1 point, with different time-independent probabilities. Thus, one option had a higher expected value than the other one (i.e., the option with a higher probability of reinforcement).

Participants

Our sample included 47 undergraduate students of Psychology from the University of Guadalajara (51% male). They were recruited by an announcement of their professor. Participants entered on a raffle, and the winner would get real money ranging from 100 to 300 Mexican pesos (approximately 5 to 15 USD) depending on the performance on the task.

Materials

Participants faced two choice tasks on a computer screen, presented in random order. They made their choices using the mouse of the computer. At the end of the experiment, participants had to fill a custom-made questionnaire about task comprehension and gambling habits. The experimental program was written using the OpenSesame software.

Procedure

Participants were randomly assigned to one of two groups: *experience* and *mixed*. Groups differed on the presentation format of the problems (see Table 10). Specifically, the *experience* group faced two unlabeled buttons, and had to choose between both options 100 times. The *mixed* group was presented with both the probabilistic description of the problems, and the experienced outcomes, as they also had to choose between them 100 times.

In each task, the participants faced a *fixed* and a *changing option* associated to two buttons whose position –left or right– was also randomized for each participant. The *fixed option* gave 1 point with a probability or 40% in both problems, while the *changing option* gave 1 point with a probability of 20% and 80% for problems A and B, respectively.

Every choice made by the participants had real consequences for them, as they got as many points as the sum of the outcomes of the 100 trials for both choice tasks (1 point = 2 Mexican pesos). Participants saw, highlighted in red, the amount of points they got in each trial, and also the amount of points they missed on the other option.

Table 10

Description of the problems for the mixed group (original in Spanish)

Problem A	**Problem B**
Option A: win 1 point with a 40% chance or win 0 points otherwise.	Option A: win 1 point with a 40% chance or win 0 points otherwise.
Option B: win 1 point with a 20% chance or win 0 points otherwise.	Option B: win 1 point with an 80% chance or win 0 points otherwise.

Results and Discussion

Figure 14 shows the average proportion of choices of the *changing option* in blocks of 25 trials for problem A (1 point with a probability of 20%) on the left panel, and problem B (1 point with a probability of 80%) on the right panel. Our prediction was confirmed, as *mixed groups* displayed consistently, across problems and blocks, a higher preference towards the optimal option (i.e., *fixed option* in problem A and *changing option* in problem B). One-tailed *t*-tests corroborated this assertion both in problem A, $t(45) = 1.94$, $p = .029$; and problem B, $t(1, 45) = 1.79$, $p = .040$.

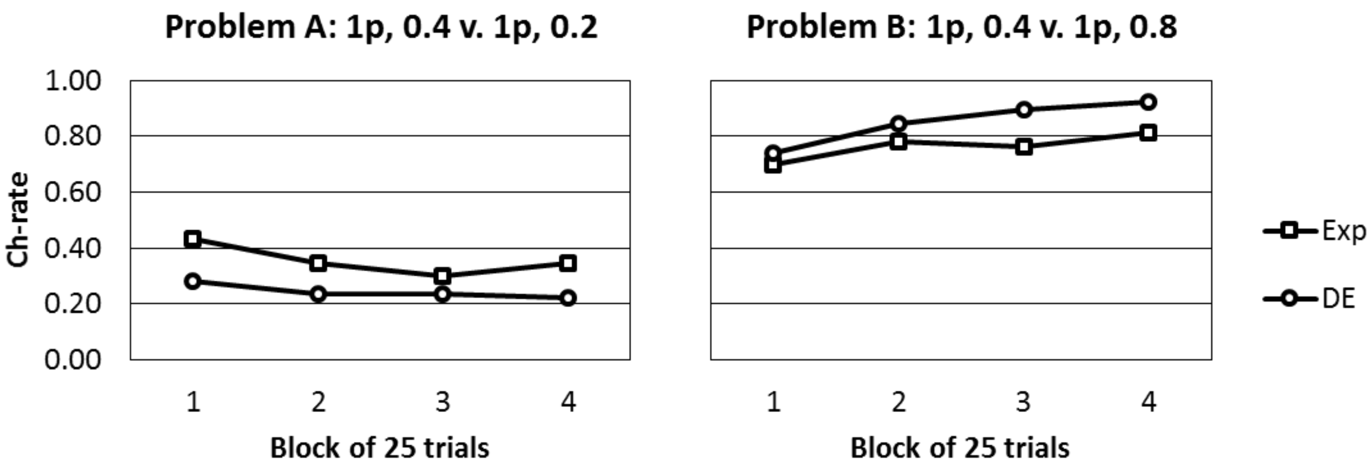

Figure 14. Choice behavior expressed as *changing option* rate in blocks of 25 trials. Left panel shows data for problem A and right panel for problem B.

Optimal behavior for tasks such as those used on the present experiment consists in evaluating which one is the best option, and sticking to it for the whole task. As the probabilities are static, the chances that a particular option will lead to reinforcement are the same on each trial. Hence, choosing invariably the *fixed* and the *changing option* in problems A and B, respectively, would maximize profit for every participant.

In the questionnaire about task comprehension and gambling habits, one third of the participants, no difference between groups, mentioned the task to be following some sort of pattern so that prizes would appear for some times in a row, maximum. Accumulated research (Tune, 1964) has shown that human participants tend to expect more short runs and less long runs in randomly generated sequences when compared to their mathematically expected distributions. In two-choice learning tasks, it has been suggested that participants may be interpreting the discrepancy between the distribution that they expected and the distribution that they encounter in every trial as if their probabilities were conditional (Beach & Swensson, 1967). This behavior has been documented since the summer of 1913 in a Monte Carlo casino, when most gamblers starting to exaggeratedly bet on red, after fifteen blacks had shown up in a roulette wheel. However, the roulette kept on landing in black eleven times more. Gambler’s behavior showed that they were assuming that the randomness of the wheel should compensate the long run and land on red (Lehrer, 2009, pp. 66). Regardless of instructional control, this pattern –the gambler’s fallacy– appears to be operating in our participants’ behavior, too. Additionally, Study 2 suggests that participants do not comprehend properly the descriptions commonly used in probability schedules.

Study 4

Method

Different concurrent variable-interval schedules were presented to the participants in a single factor, between-participants design with two levels of source of information available –only experience or experience and descriptions–. Two schedules delivered reinforcers of different magnitudes with different time-dependent probabilities. Reinforcer magnitude for each schedule was adjusted so that both options had identical expected values.

Participants

Our sample included 10 undergraduate students of Psychology from the University of Guadalajara (80% women) that were divided in two groups of 5 (Takahashi & Shimakura, 1998). They were recruited by an announcement of their professor. Participants earned virtual money in each session. At the end of the experiment, one session would be randomly drawn and its corresponding participant would get that same amount of real money, ranging approximately from 100 to 400 Mexican pesos (approximately 5 to 20 USD) depending on the execution on the task.

Materials

Participants faced the concurrent variable-intervals schedules on a computer screen, where their choices would be made using the mouse of the computer. At the end of the experiment, participants had to fill a custom-made questionnaire about task comprehension and gambling habits. The experimental program was written using the OpenSesame software.

Procedure

Each experimental session lasted 10 min, except the very first ones for some participants due to an error in the program code. Therefore, additional sessions were run by those participants on the first experimental condition to ensure data consistency. After each session, participants could take a break of 5 min if they wanted to. The median number of sessions per day, and average days that they came to the lab were 6 and 3.5, respectively. As we anticipated this large amount of work by our participants, they were allowed to use earphones and listen to music if they wanted to, following previous work that also allowed the participants to engage in other behaviors while participating in long experiments (Galizio, 1979).

Participants were randomly assigned to one of two groups: *experience* and *mixed.* Groups differed on the instructions they were given at the beginning of each session and the presentation format of the problems. variable-interval values, instructions and presentation format were adapted from the work of Takahashi and Shimakura (1998), in order to facilitate results comparison.

Instructions for *experience group* only mentioned the basic elements of the task, while instructions for *mixed group* also included reference to the functioning of the variable-interval schedules. Furthermore, the task included informative labels above the buttons in *mixed group*, reporting the average number of reinforcers that could be earned in one session, and their magnitude (see Figure 15).

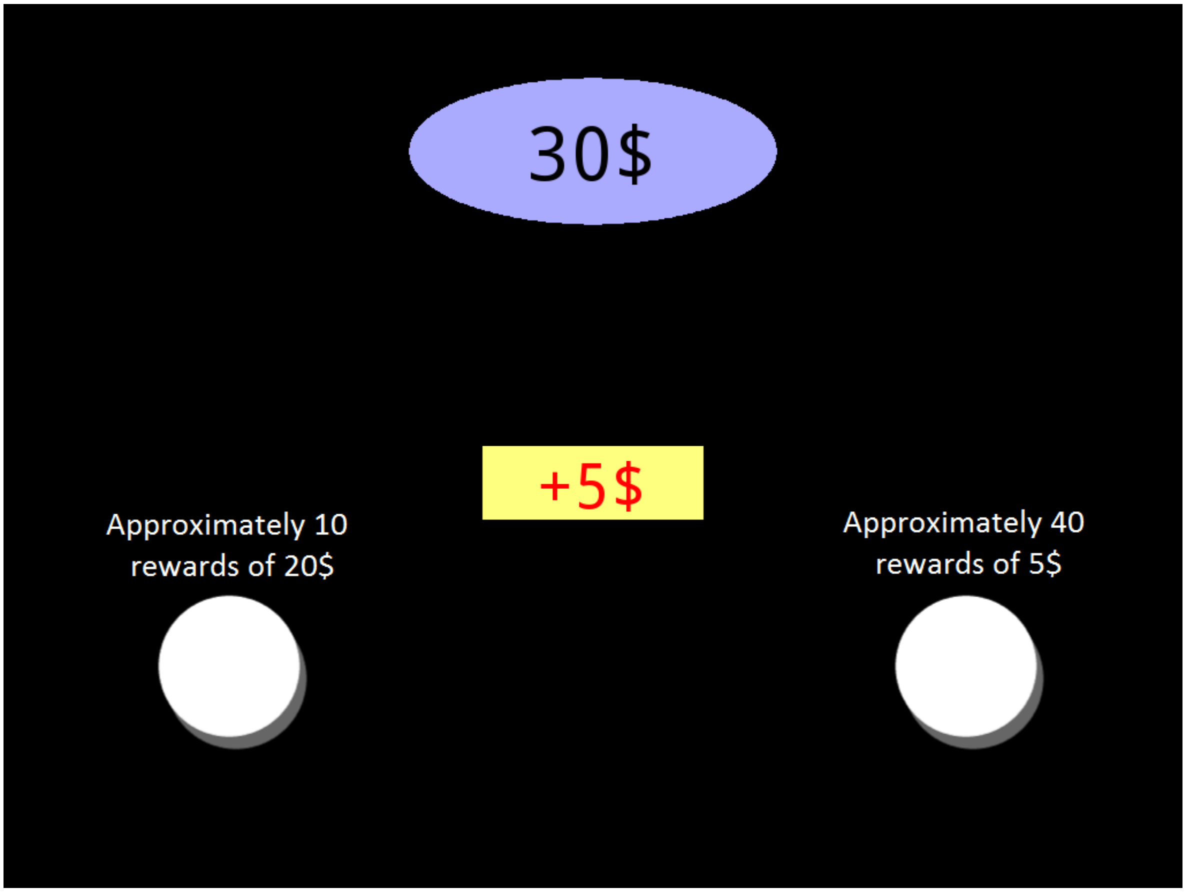

Figure 15. Screenshot of the task for the *mixed group*: a counter of the total amount earned to that point (above), the display of the reinforcer that was earned just before (center), and the two response buttons with their respective labels. These labels were not available for the *experience group.*

The instructions of the task appeared at the beginning of each session and there was no time limit for reading them before starting the task. Instructions for the *experience group* read as follows (original in Spanish):

Your task consists in increasing the counter in the superior part of the screen by clicking on the buttons of the screen. Sometimes, when you click on the buttons, you will get rewards whose amount will be displayed at the central box, and the counter will increase *accordingly. There are two buttons, and each one of them can increase the counter independently. You can choose between them freely.*

This session will last 10 minutes, and within this period you can get money. Please, get as much money as you can. Once the experiment is over, a raffle will be carried out between the participants, and the winner will get the money obtained in one of his sessions, randomly chosen.

The instructions for the *mixed group* were identical, with the exception that the following paragraphs were added between the previous paragraphs:

The number of clicks on the buttons is not relevant to increase the counter. The situation in which clicking a button will raise the money counter appears sometimes on each button. You can increase the money counter by clicking on the button in that moment. This situations appear at random intervals on each button.

You can get the rewards shown above each button by clicking on the left and right buttons. The rewards above the right button can be obtained by clicking the right button and the rewards above the left button can be obtained by clicking the left button. You can obtain these rewards during one session.

The order of the different conditions for each participant followed a quasi-complete Latin square. In each condition, the participants faced a *fixed option* on the left key, and a *changing option* on the right key. The *fixed option* provided 20 pesos (1 USD = 20 pesos, approximately) following a VI 60-s on every condition, while the *changing option* had different variable-interval and reinforcers values on each condition. *Changing option* parameters were adapted so that expected values on both options remained identical (i.e., longer variable-interval schedules delivered proportionally higher reinforcers and vice versa) (see variable-interval and reinforcers values for each condition in Table 11). A total of 30 interreinforcement intervals were arranged for each variable-interval schedule following the method of Catania and Reynolds (1968).

Participants in the *mixed group* faced one extra condition with false instructions that was not presented to the *experience group*. The purpose of this phase was to have an extra measure of the presence or absence of instructional control. On this condition, the *fixed option* remained the same as in the rest of the experiment, while the *changing option* was accompanied by a false description. The label of this option was identical to the *fixed option* VI 60-s (i.e., "about 10 rewards of 20 pesos"), while its actual schedule was a VI 180-s with a reinforcer of 20 pesos. Therefore, the label was false regarding reinforcer density: it stated that reinforcer was 3 times more frequent than it actually was. This extra condition was added to the procedure to have another measure of evidence of instructional control on the task.

Table 11

Fixed and changing option variable-interval values and reinforcer magnitudes across conditions. Only mixed group faced False Condition.

Condition	***Fixed option***	***Changing option***
1	VI 60-s (20$)	VI 7.5-s (2.5$)
2	VI 60-s (20$)	VI 15-s (5$)
3	VI 60-s (20$)	VI 60-s (20$)
4	VI 60-s (20$)	VI 180-s (60$)
5	VI 60-s (20$)	VI 480-s (160$)
False*	VI 60-s (20$)	VI 180-s (20$)

Each condition remained in effect until at least 3 sessions were run and the standard deviation of the response rate on the *fixed option* (i.e., number of responses on the *fixed option* divided by the total number of responses) on the last 3 sessions was smaller than .1.

Results and Discussion

Participants from both groups showed a similar, high response rate across the experiment, $F(1, 8) = 3.78$, $p = .088$, average response rate being 2.6 resposes / sec, even for participants from *mixed group,* whose instructions stated that the number of responses was not important to get the reinforcers. We were informed by personal communication with dr. Takahashi that participants in their experiment had moderate response rates (Takahashi & Shimakura, 1998). This contrast may have been caused by the difference between our apparatus: while they used two small levers of 2.2 x 2.0 x 0.6 cm for the participants to respond, we used standard pc mouses.

Regarding choice behavior, Figure 16 shows the ratios of the response rates in both components (i.e., average response rate on the *fixed option* divided by average response rate on the *changing option*), plotted against the ratio of the reinforcer frequencies and against the ratio of the reinforcer magnitudes, from the last 3 sessions of each condition. Similar results were obtained when the ratios of the allocation times in both components were used instead of response rates. When, in a particular session, the total number of reinforcers were obtained just on one side, this data was not considered on the data analysis (Takahashi & Shimakura, 1998). This treatment leaded to analyze 92% of the data. However, condition VI 60-s – VI 480-s from S3 was completely neglected, as the participant never obtained one single reinforcer from the *changing option* (i.e., VI 480-s) on the 3 sessions that were run.

The only participant that consistently matched her choice behavior to some property of the options was S2. She typically exhibited a higher preference towards the option that contained the greater reinforcer magnitude, $t(3) = 4.70$, $p = .018$. This pattern leaded to a sub-optimal behavior: she got less reinforcers than she would if she had been indifferent between both options, or if she had matched her behavior to reinforcer frequency (Baum, 1974; Herrnstein, 1961).

When the slopes of both groups were compared, *t*-tests revealed them to be similar, both when the independent variable was the ratio of reinforcer frequencies, $t(8) = .02$, $p =$

.985; and the ratio of reinforcer magnitudes, $t(8) = -.09$, $p = .932$. Therefore, every participant exhibited sub-optimal behavior on this experiment, as none of them matched choice with reinforcer frequency (Baum, 1974; Herrnstein, 1961). Except for S2, every participant in our experiment could appear to be insensitive to both reinforcer frequency and magnitude.

However, our interpretation of the linear models slope being not different from zero, relies on the prediction from Generalized Matching Law (Baum, 1974) that participants were sensitive to both components. Due to the properties of the variable-interval schedules that were used in our experiments, these components "cancel each other": if one option delivered reinforcers twice as frequent as another one, those reinforcers were halved in magnitude. Therefore, both options had identical expected values.

Mixed group participants' behavior on the condition with false descriptions was compared to their behavior on the condition VI 60-s – VI 60-s, in order to have another analysis of instructional control, as both conditions had identical descriptions, while different contingencies. The condition of false instructions comprised identical descriptions to the condition of VI 60-s – VI 60-s, 20 pesos in each option, while the actual contingencies were VI 60-s – VI 180-s, 20 pesos in each option (i.e., the description of the reinforcer magnitude was accurate, but the description of the reinforcment frequencies was not). Data analysis from the last 3 sessions from both conditions revealed differences on the ratio of response rates on both options, $F(1, 28) = 9.55$, $p = .004$. Therefore, participants adapted their behavior when contingencies changed, resulting in a more optimal behavior than if they had followed the false descriptions. Especifically, on the condition VI 60-s – VI 60-s, the average rate of response ratios was .96. This is, participants fairly distributed their responses between the two options. However, on the condition with false descriptions, the average rate of response ratios was 2.02. This is, participants responded twice as much to the *fixed option* (VI 60-s) than to the *changing option* (VI 180-s, described as VI 60-s).

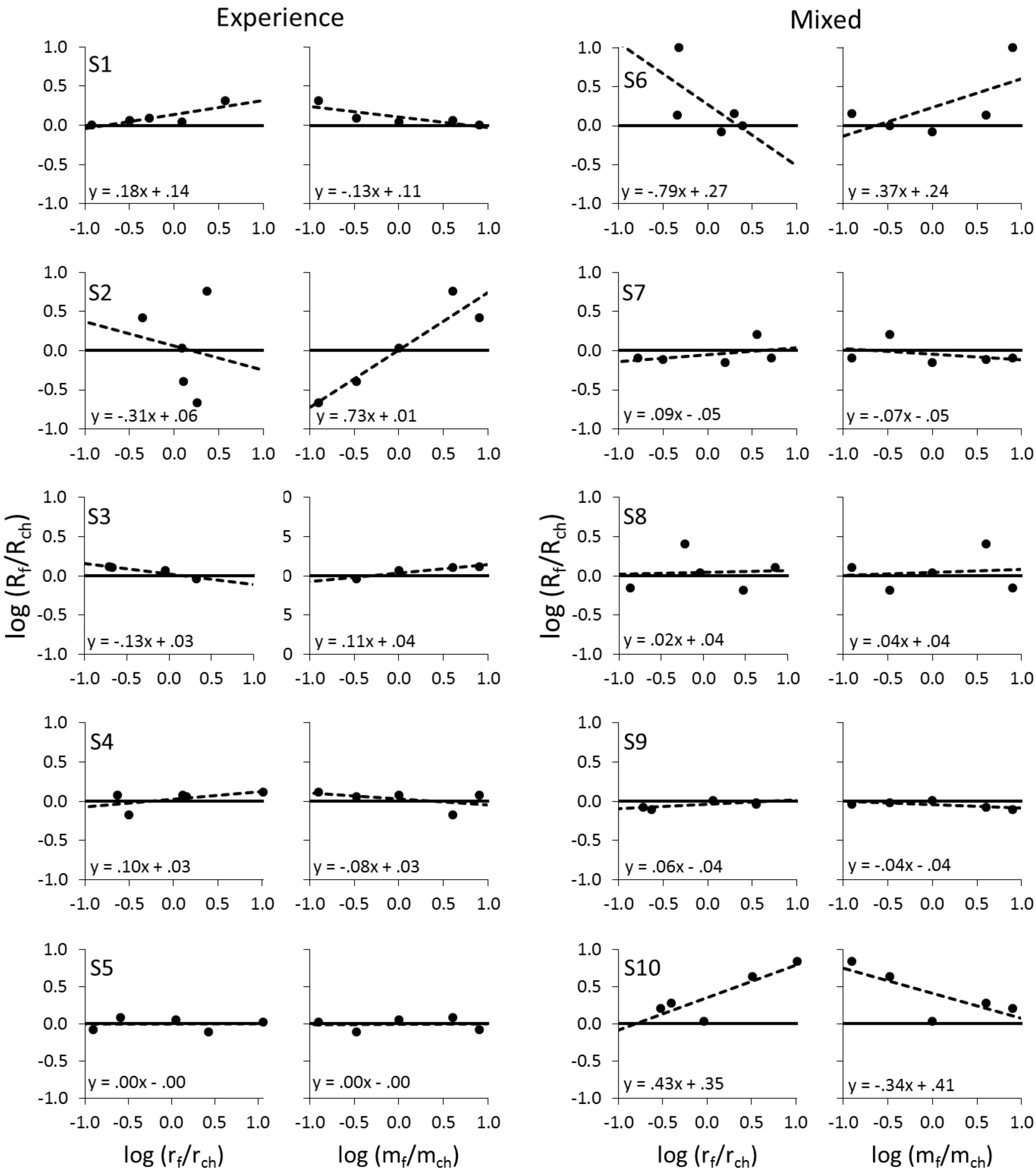

Figure 16. Ratios of response rates in the two components plotted against the ratios of reinforcer frequencies (left-hand graphs) and against the ratios of reinforcer magnitudes (right-hand graphs) in the two components using double logarithmic coordinates (base 10),

for the participants in *experience* (left panel) and *mixed* (right panel) groups. Data from the last 3 sessions of each condition were used. Solid lines represent predicted behavior by the Generalized Mathing Law, and dashed lines represent best fit linear functions. Equations for linear functions are represented at the bottom of each graph.

Therefore, our interpretation that participants were not being indifferent to the properties of the options finds support on these results too. Optimal behavior on this condition would be to respond three times as much on the *fixed option* than on the *changing option*, as their only difference was that the former delivered three times more reinforcers than the latter. Participants showed a, still sub-optimal, matching behavior to reinforcer frequencies by responding twice as much on the *fixed option* compared to the *changing option.*

As R^2 values meassure the proportion of variation of the dependent variable explained by the independent variable, and most of the obtained linear models were consistent with the null hypothesis, this meassure was not used because it gave artificially low values of model fitting (e.g., R^2 values for both regression models of S5 equaled 0, even if the model clearly fits the data). Results showed that both groups behaved similarly, and the slopes of their models were not significantly different from zero.

General discussion

Participants from Study 3 showed evidence of instructional control in a choice task using probability schedules with different expected values. *Mixed group* exhibited a more optimal behavior across blocks and problems compared to *experience group.*

On the other hand, participants from Study 4 showed no sign of instructional control when confronted with concurrent variable-interval schedules where both options had identical expected values. Most of our participants, regardless of their experimental condition, distributed their responses evenly between both options. Moreover, in the condition with false descriptions where both schedules were described likewise, they exhibited a preference towards the option with a higher expected value.

Table 12

Final evidence of instructional control as a function of expected values of the options and type of schedule

	Probability	**Variable-interval**
Similar expected values	Negative evidence	<u>Negative evidence</u>
Different expected values	<u>Positive evidence</u>	Positive evidence

These results taken together (see Table 12), point that the difference between the instructional control that was found on previous work using concurrent variable-interval schedules (Takahashi & Shimakura, 1998) and the absence of it on the experiments that were framed on the description-experience gap paradigm (Jessup et al., 2008; Lejarraga & Gonzalez, 2011) was caused by the relative expected values that were adopted.

Further interpretation of our data suggests that relative expected values seem to be important because, given the parameters used on the present experiments, it facilitated options' values discrimination. When tasks comprised reinforcers of different magnitudes, other parameters (i.e., probability of reinforcement or duration of the variable-interval schedule) were adjusted so that the expected value of each option remained similar. Therefore, discrimination of the relative expected values are important to put participants' behavior under instructional control as long as they make one option clearly better than the other one. Thus, the descriptions of the options clarify this difference and make the participants choose more optimally. On the contrary, when both options have similar expected values, adding objective information about them does not affect choice behavior.

Limitations

Only two choice tasks were used in Study 3 so the results could be compared with those studies on the description-experience gap that are fundamental for the current experiment (Jessup et al., 2008; Lejarraga & Gonzalez, 2011). Further research would be needed using additional tasks to add more validity to our analysis.

First sessions of Study 4 were shorter than 10 min. This was caused because the program required a time limit together with a trial number limit (i.e., loops, or iterations). We initially set a limit of 1000 trials which still would allow for a response rate of 1.5 responses / sec. However, due to the higher response rate of 2.6 responses / sec that our participants exhibited, this limit was overpassed on the first sessions. Therefore, it was changed to 2500 so it would not be reached before the 10 min time limit. Still, additional sessions were run to ensure data consistency before introducing a new condition.

No criterion of absence of directional trend in choice proportions was introduced in Study 4. Yet data analysis excluding conditions that could be affected by such a criterion was conducted and similar results were obtained on this experiment. Nonetheless, we encourage researchers to include it on further research.

While subjects were advised on Study 3 about the change to a new problem, this was not the case in Experiment 2 for *experience group*, following the method from Takahashi and Shimakura (1998). However, given that these participants' behavior was not different from *mixed group*, which had that information, this procedural difference seems unimportant.

Conclusion

In conclusion, the present studies showed that (1) positive evidence of instructional control was found in a choice task with probability schedules and different expected values between options; (2) negative evidence of instructional control was found in a choice task with variable-interval schedules and similar expected values between options; and (3) when taking these results together with previous research (Jessup et al., 2008; Lejarraga & Gonzalez, 2011; Takahashi & Shimakura, 1998), relative expected values was demonstrated to be a fundamental factor on understanding the presence of instructional control in choice tasks. We conclude that the relevance of this factor relies on its capacity to make participants' decisions easier by adding descriptions: participants better discriminate which is the best option.

Therefore, future lines of work shall include studying a choice task where both reinforcers magnitudes and options' expected values are different, as we would expect to find instructional control. Additionally, we have treated for the purpose of the present paper the relative magnitudes of the reinforcers as a binary factor –similar or different–, but further research could examine its contribution to instructional control in choice behavior by establishing a continuum: not every different consequences' values in the world are equally different.

Discussion

Summary of results

The meta-analysis from Study 1 evaluated the general framework of the description-experience gap in terms of the results and explanations that characterize the existing literature. Study 2 assessed the probabilistic descriptions of the options as a plausible cause of the absence of their control on participants' behavior. Finally, Studies 3 and 4 jointly evaluated the contribution of the type of schedule, and the relative expected values between options, to the existence of instructional control.

These four studies conducted within this Ph.D. thesis revealed different results that can be recapitulated as follows:

1. The meta-analysis showed that the description-experience gap is influenced by (1) as-if rare events were over- and under-weighted in description- and experience-based tasks; (2) the task domain; (3) the probability of the rare event; (4) the existence of a certain option; (5) the description task paradigm; and (6) the experience task paradigm.
2. The experimental studies showed that the presence of instructional control on the different choice tasks that were used is not influenced by (1) the demanding nature of the probabilistic descriptions of the options; or (2) the relation between probability of reinforcement and time. However, they showed that (3) it is determined by the relative expected values of the options.

Instructional control on the description-experience gap

Instructional control is an interesting function to characterize the description-experience gap from a behaviorist paradigm. The meta-analysis from Study 1 revealed that the usual methodology on description-based tasks involves probabilistic or graphic descriptions of the options and asking participants to make a choice. On the other hand, experience-based tasks are set as two buttons for the participants to make repeated choices, no other information available. Therefore, research on the description-experience gap actually compares two different functions. One of them depends on the previous learning history of the subject –description-based tasks–, specifically with the learning history of mathematical/economical instructions following behavior. On the contrary, the other function depends on the history that is formed during the task –experience-based tasks–, specifically with the engagement of the subject on concurrent probability schedules. It is no surprising that there is a “gap” on the results, as they are different things. From this argument we could deduce an interesting future experiment. People with a better history of following probabilistic instructions and choosing between them, such as economists or mathematicians, should demonstrate a higher presence of instructional control compared to the participants from Study 2.

However, we have to be careful when stating that the contingencies and instructions of a schedule are different things. They are functionally different when facing the task because they have different learning histories. But instructions following and other human verbal behavior, are not essentially different from the rest of human behavior. The same principles apply in all of them. Results obtained on instructional control of human loss avoidance (Galizio, 1979) are of great help to understand this point. When false instructions were given to the participants stating that they were under a multiple schedule with different variable-interval schedules and one extinction schedule, while no point loss was actually programmed, they continued responding according to the instructions. However, when the same instructions were given on a subsequent condition involving very short variable-intervals (i.e., participants had to respond at a high rate to avoid point loss), they stopped following instructions. Therefore, the same stimuli acquired the function of a discriminative or delta stimulus for a certain behavior depending on its consequences. When consequences were positive (i.e., no point loss), even if the instructions were false, participants continued responding as they stated. Only when negative consequences followed instruction following behavior, participants' behavior fell under the control of schedule contingencies. Another interesting possibility for further research arises here, namely, assessing the effect of false instructions on concurrent probability schedules: will instructional control still depend, as we assert, on the relative expected values?

A similar rational would apply on a situation where the experimenter, previously to implementing the actual task to the participants, creates an untrusting vs trusting situation. This has been previously done on an experiment on self-control with children using the marshmallow test. The authors had the children doing the marshmallow test on an environment that was experimentally manipulated to be either reliable or unreliable, and showed that a higher proportion of children from the former group exhibited instruction following behavior by waiting in order to get two marshmallows instead of just one, when compared with the latter group (Kidd, Palmeri, & Aslin, 2013). Therefore, it would be also interesting to evaluate instructional control as a function of immediate previous experience.

In summary, we could define the description-experience gap from a behaviorist paradigm as the difference on choice when comparing probabilistic instructions following behavior with behavior governed by the contingencies of probability schedules. Furthermore, obtained results in the empirical studies of the present work suggest that instructional control in choice tasks appears when instructions, functionally defined as discriminative stimuli, serve to discriminate a more optimal response pattern. This assertion remains true in previous literature demonstrating instructional control in response rates (Galizio, 1979; Matthews et al., 1977; Shimoff et al., 1981), and in choice behavior under concurrent probability schedules (Barron et al., 2008; Braveman & Fischer, 1968; Fantino & Esfandiari, 2002), concurrent time schedules (Hackenberg & Joker, 1994), concurrent ratio and interval schedules (Catania et al., 1982; Matthews et al., 1985), and the prisoner dilemma (Baker & Rachlin, 2001). In all of them, instructions signaled a behavior pattern to increase reinforcement frequency.

Further research has been proposed through this section, namely, (1) studying instructional control under concurrent probability schedules using participants with expertise on mathematics or economy; (2) assessing false instructions following behavior under concurrent probability schedules; and (3) evaluating instructional control as a function of immediate previous experience. The next section will serve to better understand the properties of probability schedules and their differential characteristics when compared with other resembling schedules.

Comparison of choice schedules

Studying behavior under concurrent probability schedules has lead researchers to the study of probability matching. Probability matching is a phenomenon that appears when an organism matches its choice behavior to the relative probabilities of getting reinforcement on the options, instead of maximizing income by responding on the richest alternative. It has been shown in animals using dependent schedules (Bitterman, 1965) and both independent and dependent schedules (Brunswik, 1939), and in humans using independent schedules (Fantino & Esfandiari, 2002) and dependent schedules (Shanks, 1990).

Probability schedules are defined by their outcomes an associated probabilities. However, outcomes' probabilities cannot be experienced directly. Participants engaging on repeated choices between options will experience either outcome presence or absence after each response. Therefore, probabilities are experienced as frequencies: an option that is programmed to give 10 points with a .20 probability, 0 points otherwise, will deliver the reinforcer after an average number of 5 responses. This approach to probability schedules schedule resembles a variable-ratio schedule, as a participant engaging on a VR 5 will also receive a reinforcer after an average number of 5 responses. By means of this similarity, both types of schedule have in common the definition of the average response requirement to obtain reinforcer: the average response requirement on a variable-ratio schedule VR X equals the inverse of the probability of a probability schedule *p*.

$$\text{VR X} = \frac{1}{p}$$

Table 13 shows this identification for the description-experience gap tasks from Study 2 that were identified on the present Ph.D. thesis as a source of absence of instructional control (Jessup et al., 2008; Lejarraga & Gonzalez, 2011), and from Study 3.

Table 13

Equivalence table between probabilistic and concurrent variable-ratio schedules definitions for problems from Studies 2 and 3.

	Probabilistic definition		***Concurrent***
	Option A	Option B	***variable-ratio***
Study 2 problem A	3 p (1.00)	4 p (.80), 0 p (.20)	CR 1 (3p) – VR 1.25 (4p)
Study 2 problem B	3 p (1.00)	64 p (.05), 0 p (.95)	CR 1 (3p) – VR 20 (64p)
Study 3 problem A	1 p (.40), 0 p (.60)	1 p (.20), 0 p (.80)	VR 2.5 (1p) – VR 5 (1p)
Study 3 problem B	1 p (.40), 0 p (.60)	1 p (.80), 0 p (.20)	VR 2.5 (1p) – VR 1.25 (1p)

Note. Probabilistic definitions include number of points of each outcome followed by associated probability between parentheses. Concurrent variable-ratio schedules definition include average number of responses to obtain reinforcer followed by number of points between parentheses. CR = continuous reinforcement (each response is followed by reinforcer).

This integration between probability and variable-ratio schedules facilitates design of further research on the description experience gap by means of what has been already done using variable-ratio schedules. For example, human behavior under concurrent variable-ratio schedules have been shown to maximize reinforcement by almost exclusively responding on the best alternative (Striefel, 1972; Weiner, 1966). Also, similar results have been found on scholar education applied research (Mace, 1990) and basic animal research with pigeons (Herrnstein & Loveland, 1975) and rats (MacDonall, 1988). However, human participants with problems of development exhibit sub-optimal execution (Schroeder, 1975).

Still, there are three differences between variable-ratio schedules and the probability schedules commonly used on the description-experience gap field: the probability distribution of the number of responses required, the consequences of the responses that do not lead to the main reinforcement and the value of the reinforcer.

First of all, a common way of constructing a variable-ratio schedule is creating an arithmetic progression, although it is an arbitrary decision (Ferster & Skinner, 1957, Chapter 7). In contrast, a probability schedule samples trial-by-trial whether the next response will result in reinforcer or not. Consequently, a VR 10 (1, 2, 3… 17, 18, 19) differs from a probability schedule that delivers reinforcement with a probability of .10 on the distribution of the number of responses that are required to obtain the reinforcer.

On our example, the probability of getting reinforcement under a probability schedule of .10 on the first trial is *Pr* = .10. On the second trial it decreases to *Pr* = (.90) (.10) = .09 (i.e., probability of not getting reinforcement on the first trial times probability of getting reinforcement on the second trial). On the third trial, *Pr* = $(.90)^2$ (.10) (i.e., probability of not getting reinforcement on the first and second trials times probability of getting reinforcement on the third trial). Therefore, the probability of getting reinforcement after n trials given a probability of reinforcement p under this schedule follows a geometric distribution (i.e., number of trials needed to get one success),

$$\mathrm{Pr}(n) = (1 - p)^{n-1} p$$

This geometric distribution of probability schedules leads to a right-skewness that contrasts with the uniform distribution of probabilities on variable-ratio schedules (see Figure 17).

As a result, short runs till reinforcement are more frequent in probability schedules compared to variable-ratio schedules, yet this difference can be eliminated on the construction of variable-ratio schedules: instead of arranging an arithmetic progression of numbers, researchers could construct a custom set of numbers so that the final distribution resembles that of a probability schedule. Also, another interesting possibility would be to evaluate instructional control under variable-ratio schedules. At least two experiments could be derived from the argument of our work, depending on whether both options have similar expected values or not.

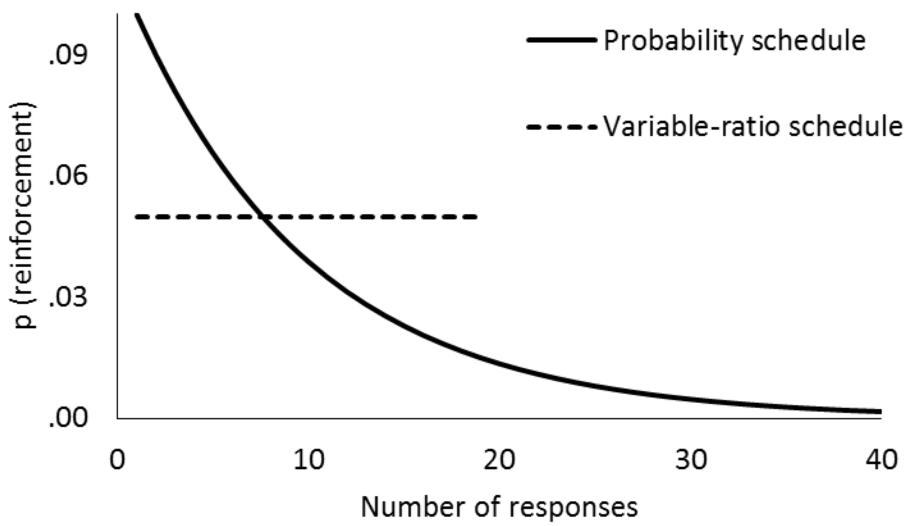

Figure 17. Probability of obtaining a reinforcer following a certain number of responses under a .10 probability and a VR 10 schedules.

A second difference refers to a type of probability schedules typically used on the description-experience gap field that were not used on the present work. Some authors (e.g., Glöckner et al., 2012) have used probability schedules that lead to two or more different amounts of reinforcer constantly, as in a schedule that delivers 10 points with a probability of .2, 4 points otherwise. Such a property could be included on variable-ratio schedules by delivering small amounts of reinforcer after each response until the criteria is met so the main reinforcer is delivered.

Finally, as the meta-analysis from Study 1 showed, even if most studies on the description-experience gap relied on positive outcomes, some authors used negative outcomes (e.g., Yechiam, Barron, & Erev, 2005) or both (e.g., Erev et al., 2010). Behavior analysts have also shown interest on the effect of punishment on concurrent schedules and the matching law (Davison & McCarthy, 1988, Chapter 8.3).

Research with pigeons under concurrent variable-interval schedules has shown that punishment of individual responses lead to proportional reduction of responding on each option (Holz, 1968). Additional research with rats demonstrated that increasing shock frequency for one option while keeping the other option constant, makes relative rate of choices to decrease on the former while increasing it on the later (Deluty, 1976; Farley & Fantino, 1978). The addition of identical variable-interval schedules of punishment to concurrent variable-interval schedule of reinforcement with unequal reinforcement frequencies moved pigeons' behavior towards overmatching (de Villiers, 1980; Farley, 1980).

Having discussed the characteristics of the schedules employed on the description-experience gap and their connection with instructional control in the light of the obtained results, the present Ph.D. thesis ends its scientific contribution. However, the absence of references to hypothetical cognitive processes while focusing on environmental and behavioral variables during the current discussion may appear gratuitous to some readers who may consider that there is no need for scientists to have a defined philosophical viewpoint. The following part will conclude about the relevance of such a philosophical definition in Psychology.

The practical implications of Philosophy

Once a particular science establishes a categorical closure (Bueno, 1995), its terms and mechanisms get objectively defined and their proper conjugation increases the boundaries of its knowledge. A set of concepts from experimental analysis of behavior have been used through this work, such as discriminative stimulus, response and reinforcer. The different combinations of these elements give their origin to reinforcement schedules. Within this framework, radical behaviorism gets to a position where it is able to explain phenomena of interest without falling in reductionisms of any way.

Radical behaviorism is radical in the sense of total, where private events are not excluded from the object of study, as in methodological behaviorism, where they are reintroduced as hypothetical constructs (Pérez, 2017b). It is a philosophy of a complete, independent Psychology: analysis of behavior. Moreover, the main mechanism underlying this paradigm is operant behavior, and functional analysis serves as the methodology to accomplish its study.

Arguing that the functional analysis is the main tool of studying human behavior is not just an academic discussion, disconnected from reality. It has practical implications when we turn the science of Psychology to its technology: the application of scientific knowledge to clinical, educational or social problems.

Applying tools that are adapted to study mechanisms from other sciences can change the proper course of Psychology. The adoption of the medical model in psychopathology offers an example of this error. By searching for “symptoms” that can be grouped into “syndromes” which are believed to be caused by a supposed dysfunctional, internal psychological mechanism, Psychology and Psychiatry professionals have turned their efforts to the creation and development, for example, of the Diagnostic and Statistical Manual of Mental Disorders (DSM). Within this framework, psychotherapy would consist in the application of specific techniques, as medication does in Medicine, developed to solve those supposed dysfunctional mechanisms. On the contrary, contextual models treat psychological problems, which are not understood as "mental disorders", as biographic and circumstantially built ways of functioning (González & Pérez, 2007).

Even the most extremes cases of psychological problems can be analyzed under the glasses of functional analysis to shed light on possible solutions. A historic case of a psychotic patient who had become completely mute and exhibiting almost no motor activity for 19 years may serve as a perfect example of the power of functional analysis. He would not even react when cigarettes were offered to him, compared to other schizophrenic patients.

The researcher eventually dropped a package of chewing gum accidentally, a situation that made the patient's eyes move toward the gum and then back to normal. The researcher realized this movement and understood that chewing gum could be used as a powerful reinforcer on this situation. Therefore, he conducted a shaping by successive approximation that involved eye movement, facial movement, lip movement, vocalizations, words, and proper verbal behavior as "gum, please" and answering questions about his name and age (Isaacs, Thomas, & Goldiamond, 1960). Functional analysis and operant conditioning demonstrated to be powerful tools to control human behavior in a situation where others would have argued that the problem was "on the subject's mind".

> *In summary, then, I am not a cognitive psychologist for several reasons. . . . We choose the wrong path at the very start when we suppose that our goal is to change the "minds and hearts of men and women" rather than the world in which they live. (Skinner, 1977, p. 10)*

Radical behaviorism's conception of human behavior as an object of study on its own terms without appealing to cognitive or physiologic processes finds justification on the three-gender ontology of philosophical materialism. The first gender of reality M_1 designates the physical, corporeal entities of the world; M_2 comprise the psychological and historical phenomena of the "inner life"; and M_3 refers to the supraindividual, objective entities that precede the individuals (Bueno, 1972). This ontological schema is translated to Psychology as body (M_1), behavior (M_2) and culture (M_3) (Pérez, 2011).

The analysis of the optical illusion of the “Necker cube” (Bueno, 2016, Chapter 4.5.3) provides an excellent example of this pluralist ontogeny. Both possible three-dimensional perceptions of the “Necker cube” constitute the M_2 (see Figure 18), but this reversible perception also involves the straight lines that are drawn in a two-dimensional space M_1 and the geometric laws that govern the structure of polyhedrons M_3.

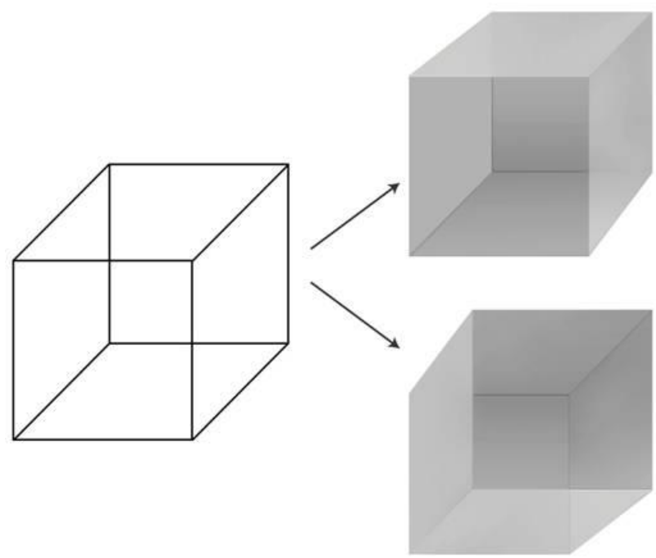

Figure 18. Three-dimensional subjective perceptions of the “Necker cube”. Adapted from “Psicología y filosofía del cubo de Necker: Para superar el dualismo y el cerebrocentrismo con el materialismo filosófico,” by M. Pérez, 2017, *Ábaco*. Copyright 2017 by Centro de Iniciativas Culturales y Estudios Económicos y Sociales.

The fundamental conclusion to be drawn here is that the psychological perception M_2 of a subject is not reduced, nor deduced from M_1 or M_3 (Pérez, 2017a). May a researcher have all the information regarding the configuration of the lines and the physiological activity of the subject (M_1) and a deep knowledge about the geometric laws of polyhedrons and the social context of the subject (M_3), he would never be able to deduce which possible three-dimensional perception of the cube (M_2) is the subject experiencing.

Depending on the particular ontogeny we adopt to see the world around us, our scientific activity can follow a path of light or darkness. The intellectual effort required to make the best choice between different paradigms is justified by the importance Philosophy has in the scientific activity. Believing that there is a possibility of making an aseptic, Philosophy-free science is nothing but an illusion.

Closing remarks

The present dissertation showed that the presence of rule-governed behavior or instructional control in choice tasks depends on the relative expected values of the options: these need to be different for instructional control to occur. For this reason, instructional control would not appear in studies conducted on the description-experience gap field comprising options with similar expected values (Jessup et al., 2008; Lejarraga & Gonzalez, 2011). Furthermore, the description-experience gap can be characterized as the difference in choice behavior of two different functions: previous learning history of probabilistic instructions following and present experience with schedule contingencies. Finally, the relevance of Philosophy as an essential, unavoidable part of our scientific activity was discussed. Looking at the description-experience gap field through behaviorist lenses involved a great integrative effort. It was well worthwhile, though.

Epilogue

Whoever saves one life, saves the world entire.

This is what the Jews saved by Oskar Schindler (played by Liam Neeson) wrote on the gold ring made out of gold fillings they gave him as a present the last time they saw each other in the movie Schindler's List (Spielberg, 1993).

In this film, Liam Neeson plays the role of a German industrialist who tried his best to save as many Jewish people as he could during the Nazi holocaust. The relationship that he establishes with his Jewish employees was initially motivated by personal profit, but through continuous personal experience he realized the extent of their suffering and decided to do something about it.

Instructional control certainly played an important part in this horrifying episode of human history. It is quite improbable that millions of Germans developed racist attitudes through personal contact with Jewish people. But instructional control, as every other behavior, depends on contextual variables. This Ph.D. thesis has made a humble contribution to the understanding of such variables.

Hence, if we want to prevent episodes like this to repeat, we should ask first:

Are we creating the appropriate context?

References

References marked with an asterisk indicate studies included in the meta-analysis.

Ader, R., & Tatum, R. (1961). Free-operant avoidance conditioning in human subjects. *Journal of the Experimental Analysis of Behavior, 4*(3), 275–276.

Álvarez, A. (2008). Sobre «Neurociencia» y Psicología. *El Catoblepas, 76*.

* Artinger, F., Fleischhut, N., Levati, M. V., & Stevens, J. R. (2012). Cooperation in a risky environment: Decisions from experience in a stochastic social dilemma. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 84–89). Austin, TX: Cognitive Science Society.

* Ashby, N. J. S. (2017). Numeracy predicts preference consistency: Deliberative search heuristics increase choice consistency for choices from description and experience. *Judgment and Decision Making, 12*(2), 128–139.

Athens, E. S., & Vollmer, T. R. (2010). An investigation of differential reinforcement of alternative behavior without extinction. *Journal of Applied Behavior Analysis, 43*(4), 569–589.

Ayllon, T., & Azrin, N. H. (1964). Reinforcement and instructions with mental patients. Journal of the Experimental Analysis of Behavior, 7(4), 327–331.

Baker, F., & Rachlin, H. (2001). Probability of reciprocation in repeated prisoner´s dilemma games. *Journal of Behavioral Decision Making, 14*, 51–67.

Baron, A., & Kaufman, A. (1966). Human, free-operant avoidance of "time out" from monetary reinforcement. *Journal of the Experimental Analysis of Behavior, 9*(5), 557–565.

Baron, A., Kaufman, A., & Stauber, K. A. (1969). Effects of instructions and reinforcement-feedback on human operant behavior maintained by fixed-interval reinforcement. *Journal of the Experimental Analysis of Behavior, 12*(5), 701–712.

* Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making, 16*(3), 215–233.

Barron, G., Leider, S., & Stack, J. (2008). The effect of safe experience on a warnings' impact: Sex, drugs, and rock-n-roll. *Organizational Behavior and Human Decision Processes, 106*(2), 125–142.

Baum, W. M. (1974). On two types of deviation from the matching law: bias and undermatching. *Journal of the Experimental Analysis of Behavior, 22*(1), 231–242.

Beach, L. R., & Swensson, R. G. (1967). Instructions about randomness and run dependency in two-choice learning. *Journal of Experimental Psychology, 75*(2), 279–282.

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica, 22*(1), 23–36.

Bitterman, M. E. (1965). Phyletic differences in learning. *American Psychologist, 20*(6), 396–410.

Blair, W. C. (1958). Measurement of observing responses in human monitoring. *Science, 128*(3318), 255–256.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Subgroup analyses. In M. Borenstein, L. V. Hedges, J. P. T. Higgins, & H. R. Rothstein (Eds.), *Introduction to Meta-Analysis*. Chichester: John Wiley & Sons, Ltd.

Borenstein, M., & Higgins, J. P. T. (2013). Meta-analysis and subgroups. *Prevention Science, 14*(2), 134–143.

Borenstein, M., Higgins, J. P. T., Rothstein, H. R., & Hedges, L. V. (2017). *I2 is not an absolute measure of heterogeneity in a meta-analysis.* Manuscript in Preparation.

Braveman, N. S., & Fischer, G. J. (1968). Instructionally induced strategy and sequential information in probability learning. *Journal of Experimental Psychology, 76*(4), 674–676.

Brunswik, E. (1939). Probability as a determiner of rat behavior. *Journal of Experimental Psychology, 25*(2), 175–197.

Bueno, G. (1972). *Ensayos materialistas.* Madrid: Taurus.

Bueno, G. (1995). *¿Qué es la ciencia?* Oviedo: Pentalfa.

Bueno, G. (2016). *El ego trascendental.* Oviedo: Pentalfa.

* Camilleri, A. R., & Newell, B. R. (2009a). The role of representation in experience-based choice. *Judgment and Decision Making, 4*(7), 518–529.

* Camilleri, A. R., & Newell, B. R. (2009b). Within-subject preference reversals in description- and experience-based choice. In N. Taatgen, H. van Rijn, L. Schomaker, & J. Nerbonne (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 449–454). Austin, TX: Cognitive Science Society.

* Camilleri, A. R., & Newell, B. R. (2011a). Description- and experience-based choice: Does equivalent information equal equivalent choice? *Acta Psychologica, 136*(3), 276–284.

* Camilleri, A. R., & Newell, B. R. (2011b). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin & Review, 18*(2), 377–384.

* Camilleri, A. R., & Newell, B. R. (2013). The long and short of it: Closing the description-experience "gap" by taking the long-run view. *Cognition, 126*(1), 54–71.

Catania, A. C., Matthews, B. A., & Shimoff, E. (1982). Instructed versus shaped human verbal behavior: Interactions with nonverbal responding. *Journal of the Experimental Analysis of Behavior, 38*(3), 233–248.

Catania, A. C., & Reynolds, G. S. (1968). A quantitative analysis of the responding maintained by interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior, 11*(3S2), 327–383.

Darwin, C. (1859). *The origin of species.* London: W. Clowes and sons.

Davison, M., & McCarthy, D. (1988). *The matching law: A research review* (2017th Ed.). New York: Lawrence Erlbaum Associates.

de Villiers, P. A. (1980). Toward a quantitative theory of punishment. *Journal of the Experimental Analysis of Behavior, 33*(I), 15–25.

Deluty, M. Z. (1976). Choice and the rate of punishment in concurrent schedules. *Journal of the Experimental Analysis of Behavior, 25*(I), 75–80.

Dews, P. B., & Morse, W. H. (1958). Some observations on an operant in human subjects and its modification by dextro amphetamine. *Journal of the Experimental Analysis of Behavior, 1*(4), 359–364.

Dutt, V., & Gonzalez, C. (2012a). Decisions from experience reduce misconceptions about climate change. *Journal of Environmental Psychology, 32*(1), 19–29.

Dutt, V., & Gonzalez, C. (2012b). Why do we want to delay actions on climate change? Effects of probability and timing of climate consequences. *Journal of Behavioral Decision Making, 25*(2), 154–164.

* Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., … Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making, 23*(1), 15–47.

Erev, I., & Roth, A. E. (2014). Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences, 111*(Supplement 3), 10818–10825.

Fantino, E., & Esfandiari, A. (2002). Probability matching: Encouraging optimal responding in humans. *Canadian Journal of Experimental Psychology, 56*(1), 58–63.

Farley, J. (1980). Reinforcement and punishment effects in concurrent schedules: A test of two models. *Journal of the Experimental Analysis of Behavior, 33*(3), 311–326.

Farley, J., & Fantino, E. (1978). The symmetrical law of effect and the matching relation in choice behavior. *Journal of the Experimental Analysis of Behavior, 29*(1), 37–60.

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement.* New York: Appleton-Century-Crofts.

Galizio, M. (1979). Contingency-shaped and rule-governed behavior: Instructional control of human loss avoidance. *Journal of the Experimental Analysis of Behavior, 31*(1), 53–70.

* Glöckner, A., Fiedler, S., Hochman, G., Ayal, S., & Hilbig, B. E. (2012). Processing differences between descriptions and experience: a comparative analysis using eye-tracking and physiological measures. *Frontiers in Psychology, 3*(173), 1–15.

* Glöckner, A., Hilbig, B. E., Henninger, F., & Fiedler, S. (2016). The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice. *Journal of Experimental Psychology: General, 145*(4), 486–508.

González, H., & Pérez, M. (2007). *La invención de trastornos mentales.* Madrid: Alianza.

* Gottlieb, D. A., Weiss, T., & Chapman, G. B. (2007). The format in which uncertainty information is presented affects decision biases. *Psychological Science, 18*(3), 240–246.

Griffin, D., & Brenner, L. (2008). Perspectives on Probability Judgment Calibration. In D. J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 177–199). Blackwell Publishing Ltd.

Hackenberg, T. D., & Joker, V. R. (1994). Instructional versus schedule control of humans' choices in situations of diminishing returns. *Journal of the Experimental Analysis of Behavior, 62*(3), 367–383.

Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia, 14*(Suppl 1), 29–37.

* Harman, J. L., & Gonzalez, C. (2015). Allais from experience: Choice consistency, rare events, and common consequences in repeated decisions. *Journal of Behavioral Decision Making, 28*(4), 369–381.

* Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making, 23*(1), 48–68.

* Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making, 21*(5), 493–518.

Hayes, S. C. (1989). *Rule-governed behavior: Cognition, contingencies, and instructional control.* New York: Springer US.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4*(3), 267–272.

Herrnstein, R. J., & Loveland, D. H. (1975). Maximizing and matching on concurrent ratio schedules. *Journal of the Experimental Analysis of Behavior, 24*(1), 107–116.

Hertwig, R. (2012). The psychology and rationality of decisions from experience. *Synthese, 187*(1), 269–292.

* Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*(8), 534–539.

* Hilbig, B. E., & Glöckner, A. (2011). Yes, they can! Appropriate weighting of small probabilities as a function of information acquisition. *Acta Psychologica, 138*(3), 390–396.

Holz, W. C. (1968). Punishment and rate of positive reinforcement. *Journal of the Experimental Analysis of Behavior, 11*(3), 285–292.

Holland, J. G. (1958). Human vigilance. *Science, 128*, 61–67.

Hothersall, D. (1997). *Historia de la Psicología* (3rd Ed.). México DF: McGraw-Hill.

Isaacs, W., Thomas, J., & Goldiamond, I. (1960). Application of operant conditioning to reinstate verbal behavior in psychotics. *Journal of Speech and Language Disorders, 25*(1), 8–12.

Jessup, R. K., Bishara, A. J., & Busemeyer, J. R. (2008). Feedback produces divergence from Prospect Theory in descriptive choice. *Psychological Science, 19*(10), 1015–1022.

* Kellen, D., Pachur, T., & Hertwig, R. (2016). How (in)variant are subjective representations of described and experienced risk and rewards? *Cognition, 157*, 126–138.

Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition, 126*(1), 109–114.

Knight, F. H. (1921). *Risk, Uncertainty and Profit.* New York: Sentry Press.

* Kudryavtsev, A., & Pavlodsky, J. (2012). Description-based and experience-based decisions: individual analysis. *Judgment and Decision Making, 7*(3), 316–331.

Lehrer, J. (2009). *How we decide.* New York: Houghton Mifflin Harcourt.

* Lejarraga, T. (2010). When experience is better than description: time delays and complexity. *Journal of Behavioral Decision Making, 23*(1), 100–116.

* Lejarraga, T., & Gonzalez, C. (2011). Effects of feedback and complexity on repeated decisions from description. *Organizational Behavior and Human Decision Processes, 116*(2), 286–295.

Lejarraga, T., Pachur, T., Frey, R., & Hertwig, R. (2016). Decisions from experience: From monetary to medical gambles. *Journal of Behavioral Decision Making, 29*(1), 67–77.

Liberman, V., & Tversky, A. (1993). On the Evaluation of Probability Judgments: Calibration, Resolution, and Monotonicity. *Psychological Bulletin, 114*(1), 162–173.

Lippman, L. G., & Meyer, M. E. (1967). Fixed interval performance as related to instructions and to subjects' verbalizations of the contingency. *Psychonomic Science, 8*(4), 135–136.

* Ludvig, E. A., & Spetch, M. L. (2011). Of Black Swans and Tossed Coins: Is the Description-Experience Gap in Risky Choice Limited to Rare Events? *PloS One, 6*(6), e20262.

MacDonall, J. S. (1988). Concurrent variable-ratio schedules: Implications for the generalized matching law. *Journal of the Experimental Analysis of Behavior, 50*(1), 55–64.

Mace, F. C. (1990). A collateral effect of reward predicted by matching theory. *Journal of Applied Behavior Analysis, 23*(2), 197–205.

* Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2017). The role of memory in distinguishing risky decisions from experience and description. *The Quarterly Journal of Experimental Psychology, 70*(10), 2048–2059.

Martens, B. K., & Houk, J. L. (1989). The application of Herrnstein's law of effect to disruptive and on-task behavior of a retarded adolescent girl. *Journal of the Experimental Analysis of Behavior, 51*(1), 17–27.

Martin, J. M., Gonzalez, C., Juvina, I., & Lebiere, C. (2014). A description-experience gap in social interactions: Information about interdependence and its effects on cooperation. *Journal of Behavioral Decision Making, 27*(4), 349–362.

Matthews, B. A., Catania, A. C., & Shimoff, E. (1985). Effects of uninstructed verbal behavior on nonverbal responding: Contingency descriptions versus performance descriptions. *Journal of the Experimental Analysis of Behavior, 43*(2), 155–164.

Matthews, B. A., Shimoff, E., Catania, A. C., & Sagvolden, T. (1977). Uninstructed human responding: Sensitivity to ratio and interval contingencies. *Journal of the Experimental Analysis of Behavior, 27*(3), 453–467.

McDowell, J. J. (1988). Matching theory in natural human environments. *The Behavior Analyst, 11*(2), 95–109.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica, 22*(3), 276–282.

Méndez, J. J. (2014). *Panfleto materialista.* Oviedo: Pentalfa.

Neef, N. A., Mace, F. C., Shea, M. C., & Shade, D. (1992). Effects of reinforcer rate and reinforcer quality on time allocation: Extensions of matching theory to educational settings. *Journal of Applied Behavior Analysis, 25*(3), 691–699.

Neef, N. A., Shade, D., & Miller, M. S. (1994). Assessing influential dimensions of reinforcers on choice in students with serious emotional disturbance. *Journal of Applied Behavior Analysis, 27*(4), 575–583.

* Oeberst, A., Haberstroh, S., & Gnambs, T. (2015). Not really the same: Computerized and real lotteries in decision making research. *Computers in Human Behavior, 44*, 250–257.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251).

Pascal, B. (2002). *Pensées.* (2002nd Ed.). Grand Rapids, MI: Christian Classics Ethereal Library.

Pérez, M. (2011). *El mito del cerebro creador: cuerpo, conducta y cultura.* Madrid: Alianza.

Pérez, M. (2017a). Psicología y Filosofía del Cubo de Necker: para superar el Dualismo y el Cerebrocentrismo con el Materialismo Filosófico. *Ábaco, 93*(3).

Pérez, M. (2017b). Psychology as a science of subject and comportment, beyond the mind and behavior. *Integrative Psychological and Behavioral Science.* Advance online publication.

* Rakow, T., Demes, K. A., & Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes, 106*(2), 168–179.

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making, 23*(1), 1–14.

* Rakow, T., & Rahim, S. B. (2010). Developmental insights into experience-based decision making. *Journal of Behavioral Decision Making, 23*(1), 69–82.

Ribes-Iñesta, E. (2000). Instructions, rules, and abstraction: A misconstrued relation. *Behavior and Philosophy, 28*, 41–55.

Russell, B. (1910). Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society, 11*(New series), 108–128.

Ryle, G. (1949). *The concept of mind* (2009th Ed.). New York: Taylor & Francis Group.

Schroeder, S. R. (1975). Perseveration in concurrent performances by the developmentally retarded. *The Psychological Record, 25*, 51–64.

Shanks, D. R. (1990). Connectionism and the learning of probabilistic concepts. *The Quarterly Journal of Experimental Psychology, 42*(2), 209–237.

Shimoff, E., Catania, A. C., & Matthews, B. A. (1981). Uninstructed human responding: Sensitivity of low-rate performance to schedule contingencies. *Journal of the Experimental Analysis of Behavior, 36*(2), 207–220.

Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1953). *Science and human behavior* (2014th Ed.). Pearson Education.

Skinner, B. F. (1957). *Verbal behavior* (1992nd Ed.). Acton, MA: Prentice Hall.

Skinner, B. F. (1969). *Contingencies of reinforcement* (2013th Ed.). New York: Appleton-Century-Crofts.

Skinner, B. F. (1977). Why I am not a cognitive psychologist. *Behaviorism, 5*(2), 1–10.

Spielberg, S. (1993). *Schindler's list*. USA: Universal Pictures.

Striefel, S. (1972). Timeout and concurrent fixed-ratio schedules with human subjects. *Journal of the Experimental Analysis of Behavior, 17*(2), 213–219.

Takahashi, M., & Shimakura, T. (1998). The effects of instructions on human matching. *The Psychological Record, 48*(1), 171–181.

Thorndike, E. L. (1898). Some experiments on animal intelligence. *Science, 7*(181), 818–824.

Thorndike, E. L. (1905). *The elements of Psychology*. New York: Mason-Henry Press.

Thorndike, E. L. (1911). *Animal intelligence*. New York: Macmillan.

Tortosa, F., & Civera, C. (2006). *Historia de la Psicología*. Madrid: McGraw-Hill.

Tune, G. S. (1964). Response preferences: a review of some relevant literature. *Psychological Bulletin, 61*(4), 286–302.

Turner, L. H., & Solomon, R. L. (1962). Human traumatic avoidance learning: Theory and experiments on the operant-respondent distinction and failures to learn. *Psychological Monographs: General and Applied, 76*(40), Whole No. 559.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323.

* Ungemach, C., Chater, N., & Stewart, N. (2009). Are Probabilities Overweighted or Underweighted When Rare Outcomes Are Experienced (Rarely)? *Psychological Science, 20*(4), 473–479.

* Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review, 111*(2), 430–445.

Weiner, H. (1962). Some effects of response cost upon human operant behavior. *Journal of the Experimental Analysis of Behavior, 5*(2), 201–208.

Weiner, H. (1966). Preference and switching under ratio contingencies with humans. *Psychological Reports, 18*, 239–246.

Wulff, D. U., Canseco, M. M., & Hertwig, R. (2017). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin.* Advance online publication.

Wulff, D. U., Hills, T. T., & Hertwig, R. (2015). Online product reviews and the description-experience gap. *Journal of Behavioral Decision Making, 28*(3), 214–223.

* Yechiam, E., Barron, G., & Erev, I. (2005). The role of personal experience in contributing to different patterns of response to rare terrorist attacks. *Journal of Conflict Resolution, 49*(3), 430–439.

* Yoon, S., Vo, K., & Venkatraman, V. (2017). Variability in decision strategies across description-based and experience-based decision making. *Journal of Behavioral Decision Making*.

Zemeckis, R. (2000). *Cast away*. USA: Twentieth Century Fox.