Determinants of the Context Dependency of Choices

Inauguraldissertation
zur
Erlangung der Würde
eines Doktors der Philosophie
vorgelegt der
Fakultät für Psychologie
der Universität Basel

von

Mikhail Sergeevic Spektor

aus Leningrad (heute St. Petersburg), Russland

Basel, 2017

UNI
BASEL

Genehmigt von der Fakultät für Psychologie

auf Antrag von

Prof. Dr. Sebastian Gluth
Prof. Dr. David Kellen

Basel, den _____

_____
Prof. Dr. Roselind Lieb

UNI
BASEL

Declaration of Authorship

I, Mikhail Sergeevic Spektor (born September 19, 1989, in Leningrad, RSFSR, now St. Petersburg, Russian Federation), hereby declare the following:

(a) My cumulative dissertation is based on three manuscripts (Gluth*, Spektor*, & Rieskamp, 2017; Spektor, Gluth, Fontanesi, & Rieskamp, 2017; Spektor, Kellen, & Hotaling, 2017). Asterisks denote equal contributions and are omitted in the main text. I contributed independently and substantially to all manuscripts in this dissertation in the following ways:

- Spektor, Gluth, Fontanesi, and Rieskamp (2017): Primarily responsible for development of cognitive model, experimental paradigm and design, data collection and analyses, and writing of the paper.
- Spektor, Kellen, and Hotaling (2017): Primarily responsible for the experimental design, data analyses, and writing of the paper. Jointly responsible for the idea.
- Gluth*, Spektor*, and Rieskamp (2017): Primarily responsible for experimental design, data collection, and eye-movement analyses. Jointly responsible for the idea and writing of the paper.

(b) I only used the resources indicated.

(c) I marked all the citations.

Basel, December 21, 2017

Abstract

Determinants of the Context Dependency of Choices

by Mikhail Sergeevic Spektor

One of the most fundamental assumptions of axiomatic economic decision-making theories is the notion of *independence*, according to which individuals evaluate choice options in isolation of each other. In other words, the presence of one option should not affect the value of other options. Decades of research accumulated a whole body of evidence that this assumption is systematically violated, resulting in so-called *context effects*. A triad comprising the similarity, attraction, and compromise effect has attracted the most interest so far, became a benchmark for multi-alternative decision-making models, and has been regarded as fundamental to decision making. Nevertheless, these context effects' universality has been challenged by identifying various boundary conditions, for example, desirability of the choice set. However, one important moderator variable that has not been systematically explored so far is *presentation format*. This is particularly important in light of recent observations that presentation format can have a substantial influence on decision making. For example, elicited risk attitudes change substantially when decisions are based on experiences and not on descriptions. In my dissertation, I aim to systematically explore how presentation format moderates context effects and which cognitive mechanisms underlie this a change of behavior. In Spektor, Gluth, Fontanesi, and Rieskamp (2017), we used an experience-based paradigm to assess the occurrence of context effects in this setting and thereby test a novel learning model. This model assumes that salient outcomes receive more attention and are thus perceived as more attractive. In line with our model's predictions, we observed the similarity effect and reversals of the compromise and attraction effects. Another recent promising advance is the use of perceptual decision-making tasks as proxies for preferential choice. In Spektor, Kellen, and Hotaling (2017), we adapted one such popular task, the rectangle-size task, to investigate the boundary conditions of the attraction effect and the existence of repulsion effects. We observed that the arrangement of stimuli on-screen had a substantially stronger influence on choices than stimulus design. Using a somewhat similar approach, in Gluth, Spektor, and Rieskamp (2017), we used a preferential task with perceptually coded features to investigate an apparent inconsistency regarding the influence of a third option's value on relative choice accuracy between the other two options. We found that the third option's value did not have this influence. However, without time pressure, we observed classical context effects that disappeared under time pressure. We found that value-based attentional capture provided a coherent account of the data in the experiments. I conclude that context effects of preferential choice are highly dependent on the presentation format and argue that attention plays a crucial role for this dependency. Future research should investigate such an attentional explanation of the presentation-format dependency of context effects.

Acknowledgements

I am deeply grateful to all persons who accompanied me throughout my Ph.D. and made it an experience I wouldn't want to miss.

In particular, I would like to thank my two Ph.D. advisors, Jörg Rieskamp and Sebastian Gluth, for guiding and supporting me in these years. I thank my other collaborators: Laura Fontanesi, Jared Hotaling, David Kellen, Karl Christoph Klauer, and Dirk Wulff. It has been (and continues to be) more than a pleasure working with you. I also thank Anita Todd for editing Manuscript 1 and all student research assistants who helped managing the studies and collecting data.

A huge thanks goes to my former office mates Marin and Oliver for all the good times we had, the PhD students at the Center for Economic Psychology: Laura, Janine, Peter, Rebecca, Regina, and Sebastian, and the other members of the group.

To all my new friends gained in the last years and, of course, my old friends: Thanks for being there for me.

Last but not least, I would like to thank my parents and my sister for their unconditional love and support.

Contents

1 Introduction

In our every-day lives, we are faced with many different decisions, such as what to eat for lunch, where to go on vacation, which job offer to take, whom to marry, and so forth. In all of these situations, decision makers might take situational factors into account. Think of Marina who is still indecisive whether or not to have lunch at the Thai take-away restaurant she eats at almost every day. To make her decision, she might consider that she had Thai food for the past two days in a row (in which she did not enjoy it as much as usually) or that she heard about food poisoning that a colleague got at another Thai restaurant close-by. These aforementioned cases are examples of *context-dependent preferences*: Marina's preferences are not invariant across different situations, but rather highly dependent on it.

Context-dependent preferences pose a serious theoretical challenge for axiomatic or rational economic decision-making theories, as they violate the pervasive *independence axiom* (Luce, 1959). It is the notion that the valuation of one option relative to another is independent of the other options. Despite the fact that *context effects*—intra-individually systematic preference changes depending on situational factors and therefore violations of independence—have been discussed in the literature for over half a century (Debreu, 1960; Tversky, 1972), rigorous research on factors modulating the occurrence of context effects has gained prominence only rather recently (e.g., Frederick, Lee, & Baskin, 2014). Different presentation formats seem to influence both decision making in general (e.g., Hertwig, Barron, Weber, & Erev, 2004) as well as the occurrence of context effects, sometimes even leading to their reversal (e.g., Trueblood & Pettibone, 2017).

For the most part, however, it remains elusive *how* presentation format interacts with context-dependent preferences. Which cognitive processes are at play? Under which circumstances can well-established context effects disappear or even reverse? The goal of this cumulative dissertation is to provide answers to both questions by exploring some aspects of this double-dependency of choices on context and presentation format, and provide an attention-based explanation for the observed phenomena.

1.1 Context Effects of Preferential Choice

Axiomatic economic choice theories (e.g., Savage, 1954; von Neumann & Morgenstern, 1947) are often labelled as being *normative* in the sense that if individuals were to follow their principles, they would be sure to maximize some form of desirable output (in the long run). Consequentially, these theories have served as benchmarks for human behavior ever since (e.g., Erev, Ert, Plonsky, Cohen, & Cohen, 2017; Lichtenstein & Slovic, 1971; Tversky, 1969). Typical axiomatic theories require individuals to be consistent, as formalized with a set of consistency principles (e.g., von Neumann & Morgenstern, 1947; see Rieskamp, Busemeyer, & Mellers, 2006, for an overview). One of the core axioms underlying these theories is *independence*. The independence axiom states that the relative binary choice proportions are independent of the choice set that the options are presented in (Luce, 1959, p. 9). Intuitively, this principle seems plausible: Other alternatives should be irrelevant when one evaluates two goods relative to each other. However, independence sometimes also makes less plausible predictions.

For example, let us consider a "fruit example": Imagine Elena who, when faced with a choice between an apple and an orange, picks an apple in 60% of the cases. The resulting relative binary choice proportion is thus $\frac{Pr(\text{apple})}{Pr(\text{orange})} = \frac{.60}{.40} = 1.5$. According to the independence axiom, she would still prefer an apple 1.5 times as often as an orange if the set she chooses from also comprises a pear, a melon, spaghetti, or anything else conceivable. Debreu (1960) provided a thought experiment on how the independence axiom would lead to surprising predictions; Adapted to our fruit example, let us assume that the orange in Elena's choice set came from Spain and that she is indifferent with respect to the origin of oranges. Elena likes the Italian as much as the Spanish ones. Therefore, when faced with a decision between an Italian and a Spanish orange, she would pick both equally often, and when faced with a decision between an apple and an Italian orange, she would still prefer the apple in 60% of the cases. If Elena complies with the independence axiom and decides between an apple, a Spanish orange, and an Italian orange, she needs to maintain these relative choice proportions $\frac{Pr(\text{apple})}{Pr(\text{Spanish orange})} = \frac{Pr(\text{apple})}{Pr(\text{Italian orange})} = 1.5$ and $\frac{Pr(\text{Spanish orange})}{Pr(\text{Italian orange})} = 1$. She would therefore choose the apple in $\frac{3}{7}$ of the cases and each of the oranges in $\frac{2}{7}$ of the cases. Such a choice pattern would mean that Elena now suddenly prefers oranges to apples, as $\left(2 \times \frac{2}{7}\right) > \frac{3}{7}$, even though one would intuitively expect Elena to choose the apple in 60% of the cases still. Debreu's (1960) criticism of this prediction of Luce's choice axiom (Luce, 1959) was later supported empirically, became known as the *similarity effect* (Tversky, 1972), and was established as the first context effect of preferential choice. A well-fitting verbal description of the similarity effect is that adding an option to a choice set "hurts" similar alternatives more than dissimilar ones (Tversky, 1972, p. 283).

Huber, Payne, and Puto (1982) argued later that similarity is not the only aspect determining which options are "hurt" by adding a new option to a choice set. Based on a *dominance* mechanism, they expected that new options can

boost similar options' choice proportions rather than hurt them. An option dominates another option if they both share the same attributes, the former is strictly better on at least one of these attributes while at the same time being at least equally good on all other attributes. Huber et al. (1982) have shown that adding a third, *asymmetrically dominated* option (i.e., an option that is dominated by one of the original options but not the other) to the choice set *increases* rather than *decreases* that similar option's choice share. Applied to the fruit example, instead of offering Elena to choose from an apple, a Spanish orange, and an Italian orange, she might get the opportunity to choose between an apple, an orange, and another orange that is *slightly* less sweet than the original one. According to the predictions of the attraction effect, Elena might now choose the apple, orange, and inferior orange in 50%, 45%, and 5%, respectively, of the cases—an *attraction effect*.[1]

As a potential explanatory mechanism behind the attraction effect, Simonson (1989) developed a verbal theory according to which humans seek reasons in order to justify their decisions. In our example, the slightly-less-sweet orange gives Elena a hard-to-argue-against reason to pick the slightly-sweeter original orange. Based on this theory, Simonson (1989) predicted the occurrence of a new phenomenon, the *compromise effect*. The compromise effect is supposed to arise when a newly added option renders one of the original options a compromise between the original alternative option and the new one. Again, referring back to our fruit example, imagine that Elena trades off the sweetness and juiciness of fruits when deciding between them, where the apple wins in terms of sweetness and the orange in terms of juiciness. In a choice situation involving an apple, orange, and in addition a pomegranate that is both juicier and less sweet than the orange, her choice proportions might thus be, in that order, 40%, 30%, and 30%, again violating independence.

The triad of similarity effect, attraction effect, and compromise effect has been studied the most so far (for an overview, see Rieskamp et al., 2006) and became a benchmark for multi-alternative decision-making models (e.g., Trueblood, Brown, & Heathcote, 2014). Nevertheless, other context effects have been demonstrated in the literature as well. Among them, the *phantom-decoy effect* (e.g., Doyle, O'Connor, Reynolds, & Bottomley, 1999; Trueblood & Pettibone, 2017) is most noteworthy within the scope of this dissertation. In principle, a phantom-decoy effect is a systematic change of preferences when a *phantom* option is added to a choice set, in other words, an option that is at some point in time declared non-available to the decision maker. Most often, a phantom decoy is used that asymmetrically dominates one of the available options. Such a placement has been shown to increase the relative choice share of the dominated-but-available option (e.g., Pettibone & Wedell, 2000).

A core prerequisite for all these context effects to arise is a *multi-attribute* representation of options, such as sweetness and juiciness or quality and price. A schematic illustration of option placements in a two-dimensional attribute space

[1] Note that such a preference shift would also violate the *regularity* principle which I will not discuss further in this dissertation (but see Rieskamp et al., 2006).

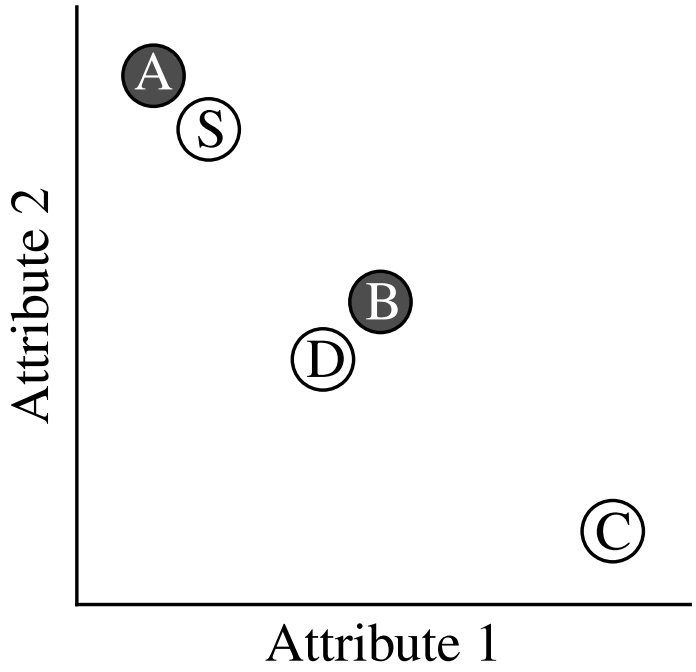


FIGURE 1.1: Typical context-effect arrangement of options in a two-dimensional attribute space. Options A and B are the core options. Option S is a similarity-effect decoy for B. Option D is an attraction-effect decoy for B. Option C is a compromise-effect decoy for B.

(the most common case) is depicted in Figure 1.1. There, options A and B form the two *core options* and options S, D, and C each form a *decoy*. In other words, the relative choice proportion of A and B when choosing from the set {A, B} changes in case a decoy is present as well. All decoy placements depicted in Figure 1.1 would lead to an increased choice proportion of B relative to A, making B the so-called *target* and A the *competitor*. The reasons why the choice share of B relative to A should increase are the following:

- When choosing from the set {A, B, S}, the addition of S "hurts" the more similar option A than the dissimilar option B (similarity effect).
- When choosing from the set {A, B, D}, the attraction-effect decoy D is dominated by B (both attributes are worse than B's), but not by A (A is better than D on Attribute 2, but D is better than A on Attribute 1; attraction effect).
- When choosing from the set {A, B, C}, the option C renders option B now a compromise between A and C. A is the best option on Attribute 2, but the worst one on Attribute 1 and vice-versa for C (compromise effect).

Using such a two-dimensional set-up, context effects have be studied in various domains. Most frequently, they have been demonstrated in consumer-choice settings (e.g., Berkowitsch, Scheibehenne, & Rieskamp, 2014; Dhar & Simonson,

2003; Doyle et al., 1999; Huber et al., 1982; Noguchi & Stewart, 2014; Simonson & Tversky, 1992), but they were also reported in decision making under risk (e.g., Chung et al., 2017; Herne, 1999; Hu & Yu, 2014; Mohr, Heekeren, & Rieskamp, 2017; Soltani, De Martino, & Camerer, 2012; Tversky, 1972; Wedell, 1991), inter-temporal choices (Gluth, Hotaling, & Rieskamp, 2017), inference tasks (e.g., Liew, Howe, & Little, 2016; Trueblood, 2012), and perceptual tasks (e.g., Choplin & Hummel, 2005; Trueblood, Brown, & Heathcote, 2015; Trueblood, Brown, Heathcote, & Busemeyer, 2013; Trueblood & Pettibone, 2017). Context effects have been found in other mammalian and non-mammalian animals (Parrish, Evans, & Beran, 2015; Scarpi, 2011; Tan et al., 2015), leading some researchers to argue that context effects are fundamental to decision making (Trueblood et al., 2013) due to their seemingly ubiquitous occurrence.

1.2 Models Explaining Context Effects

The manifold demonstrations of context effects across domains and species highlight their importance for the development of comprehensive theories of decision making. Hence, a plethora of different models has been proposed. A particularly fruitful line of models emerged within the *evidence-accumulation framework* (see Ratcliff & Smith, 2004, for a general overview of non-context-effect specific evidence accumulation models). Within this framework, it is assumed that individuals accumulate some kind of noisy evidence in favor of or against each option until a pre-defined threshold is reached. Once a threshold is reached, the decision associated with that threshold is executed. For example, in a binary lottery task, one threshold could correspond to the "risky" option and the other to the "safe" option. If the risky-option threshold is reached, then the individual chooses that option. An example of two decisions with two options and full feed-forward inhibition (i.e., evidence accumulated in favor of one option is at the same time evidence against the other, such as in the diffusion decision model; Ratcliff, 1978) is illustrated in Figure 1.2.

Being the first model to simultaneously account for all three major context effects, the *multialternative decision field theory* (MDFT; Roe, Busemeyer, & Townsend, 2001) assumes two core processes that give rise to context effects: first, attentional switching between different attributes and second, distance-dependent lateral inhibition.[2] MDFT assumes that attention randomly fluctuates between the attributes through the time-course of a decision. For example, one moment in time, the decision maker might pay attention to the quality of a product, and another moment, she would pay attention to the economy, and so on. On each such step, she would compare the options relative to each other on the attended attribute dimension. This attention-switching mechanism gives rise to the similarity effect, as similar options jointly gain and lose as their stronger and weaker attributes, respectively, are attended to, leading to a stronger competition between these options. The lateral inhibition mechanism

[2]In a nutshell, lateral inhibition means that accumulated evidence for one option suppresses the activation of all other options as a decreasing function of distance.

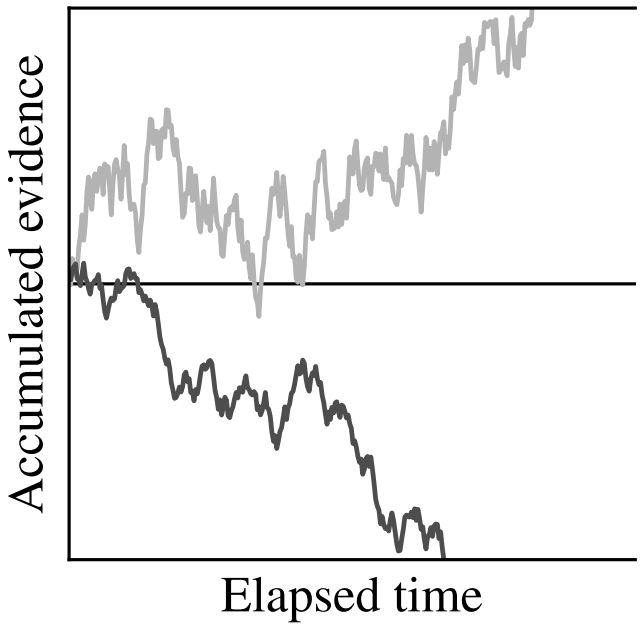


FIGURE 1.2: Illustration of an evidence-accumulation process. Depicted are two trajectories of evidence accumulation with one (light gray) reaching the upper decision threshold and the other (dark gray) reaching the lower decision threshold.

is mainly responsible for the attraction effect and compromise effect. In case of the attraction effect, the decoy will lose every comparison with the target, ultimately leading to a negative activation of the decoy. Through that negative activation and closeness of the decoy to the target, the negative inhibitory link actually *boosts* the activation of the target. The compromise effect arises for similar reasons, but is more dependent on the moment-to-moment fluctuations of valence (see Roe et al., 2001, pp. 382–384, for details).

Other evidence-accumulation models explain context effects using alternative processes, such as disproportionate weighting of negative attribute comparisons to positive ones (Usher & McClelland, 2004) or a distorted mapping of objective quantities to psychological magnitudes (Trueblood et al., 2014). Some models assume the very same attentional-switching mechanism as in MDFT (Bhatia, 2013; Usher & McClelland, 2004), others re-formulate that mechanism in terms of attention or attribute weights (Trueblood et al., 2014; Wollschläger & Diederich, 2012). However, the central role of attention fluctuating between attributes is shared among almost all of the models.[3]

Due to the wealth of models presented in the literature, efforts have been made to evaluate them on empirical grounds (e.g., Trueblood et al., 2014; Tsetsos, Usher, & Chater, 2010). However, these efforts have been bounded by the availability of closed-form solutions of model formulae (e.g., in case of the leaky, competing accumulator model, Usher & McClelland, 2004, or the MDFT, Roe

[3]It should be noted that models outside of the evidence-accumulation framework do not necessarily rely on attention-switching mechanisms (e.g., Soltani et al., 2012).

et al., 2001, but see Hancock, Hess, and Choudhury, 2018, or Berkowitsch et al., 2014, for derivations of closed-form solutions for the MDFT) or under-specifications of certain model aspects (e.g., Hotaling, Busemeyer, & Li, 2010; Tsetsos et al., 2010, in case of the MDFT). Therefore, quantitative model comparisons are sparse (e.g., Berkowitsch et al., 2014; Trueblood et al., 2014). Fortunately, qualitative model comparisons based on the patterns observed in behavior are feasible since the models make a-priori different predictions (Bhatia, 2013; Soltani et al., 2012; Tsetsos, Chater, & Usher, 2015; Tsetsos et al., 2010). The property that allows for such comparisons is the core notion of a clear representation of the options in the multi-dimensional attribute space that all models assume. For example, minor shifts in a two-dimensional attribute space determine whether the MDFT predicts a similarity effect or, if an option is only slightly worse, an attraction effect (see Tsetsos et al., 2010, Figure 5). Consequently, shifting options around in that attribute space will have a strong influence on model predictions.

1.3 Boundary Conditions of Context Effects

The various cognitive models that explain context effects make strong predictions about the boundary conditions under which context effects occur. For example, with respect to the influence of attribute distance between the target and decoy in the multi-dimensional space. Some models predict a step-like influence (Bhatia, 2013; Roe et al., 2001), others a smoothly decreasing influence of the decoy (Tsetsos et al., 2010), yet others a curvilinear influence (Trueblood et al., 2014). Despite the existence of qualitatively different predictions that would allow to create critical tests for the models, these predictions have not been evaluated thus far for the classic three context effects (but see Trueblood & Pettibone, 2017, for such an investigation of the phantom-decoy effect using a perceptual task). A boundary condition related to attribute distance that has recently been identified is the existence of prior trade-offs (or difficulty), in other words, whether individuals have a strong preference towards one of the core options before a decoy is added. It has been shown that such an initial preference attenuates the attraction effect (Farmer, Warren, El-Deredy, & Howes, 2017; Huber, Payne, & Puto, 2014) and can even lead to different preference changes depending on the target option (Evangelidis, Levav, & Simonson, 2017).

One of the most important yardsticks that model developers used when constructing their models is the ability to simultaneously predict all three context effects with one set of parameters or, more generally speaking, the co-occurrence of several context effects. So far, three studies assessed the degree of co-occurrences between context effects (Berkowitsch et al., 2014; Liew et al., 2016; Trueblood et al., 2015). All three studies, using consumer-choice, inference, and perceptual tasks, found that the context effects interact with each other: They observed a strong positive correlation between the attraction and compromise effect but negative correlations between both of them and the similarity effect. This corroborates most of the models' mechanisms, as they assume

two separate processes: One giving rise to the attraction and compromise effects (but counter-acting the similarity effect) and the other one explaining the similarity effect (usually attention shifting). More important, Trueblood et al. (2015) rarely observed all three context effects within one individual, and Liew et al. (2016) did not find any cluster of individuals that showed all three effects, even though on the aggregate level, all three context effects seemed to arise.

Huber et al. (2014) identified important boundary conditions of the attraction effect, such as the clear perception of dominance (pp. 522–523) or having to choose from unattractive choice sets (see Malkoc, Hedgcock, & Hoeffler, 2013). Under the assumption that individuals need deliberation time to form preferences, limiting the deliberation time should attenuate context effects. And indeed, time pressure has been found to decrease the magnitude of attraction and compromise effects (Pettibone, 2012) and the similarity effect (Trueblood et al., 2014).

In a recent large-scale replication attempt of the attraction effect, Frederick et al. (2014) concluded that the effect is only replicable in situations in which the options are represented by highly stylized stimuli (i.e., stimuli whose attributes are represented by numbers); Other presentation formats (e.g., wheels of fortune or photographs of fruits) did not lead to reliable attraction effects. Interestingly, the *repulsion effect* (i.e., reversed attraction effect) was found to occur just as often (Frederick et al., 2014, p. 488). Whether individuals engage in a preferential-choice or a perceptual task seems to have an influence on context effects as well. For example, Trueblood et al. (2013) used a rectangle-size task to investigate context effects using between-subject experiments. They found no compromise effect and, compared to context-effect investigations in the consumer-choice domain (e.g., Berkowitsch et al., 2014), an attenuated attraction effect. Using the very same task to investigate the phantom-decoy effect, Trueblood and Pettibone (2017) found the phantom-decoy effect to reverse in their perceptual task.

Even though many moderating variables and boundary conditions of context effects have been identified in the literature, they are connected rather loosely so far. It remains unclear why in some cases context effects flip when the presentation format changes (Trueblood & Pettibone, 2017), why a context effect as well-replicated as the attraction effect is so fragile to presentation format (Frederick et al., 2014), and, more important, which cognitive processes are responsible for these changes of behavior.

Therefore, in this dissertation, I aimed to extend our understanding of the boundary conditions of context effects:

- The goal of *Manuscript 1: "How Similarity Between Choice Options Affects Decisions From Experience: The Accentuation of Differences Model"* was to answer the cognitive-process question posed in the last paragraph: How necessary is the ubiquitous requirement of clear multi-dimensional representations of options' attributes for context effects to emerge? To do so, we chose a presentation format in which individuals have to infer options' underlying properties from trial-and-error learning. We argue

that people do not construct a clear-cut multi-attribute representation in learning settings like these, thus allowing us to assess whether such a representation poses a boundary condition for context effects.

- In *Manuscript 2: "When the Good Looks Bad: An Experimental Exploration of the Repulsion Effect"*, we wanted to explore the boundary conditions of the attraction effect, thus bridging the gap between decades of research on the attraction effect (see Simonson, 2014) and the failure to replicate it in recent large-scale replication attempts (Frederick et al., 2014; Yang & Lynn, 2014). In addition, we aimed to provide a first systematic investigation of how attribute distance between the target and decoy affect the magnitude of the attraction effect. To reach these goals, we exploited two advantages that the presentation format of perceptual decision-making tasks offers in comparison to preferential-choice paradigms: the fine-grained control over experimental stimuli and the ability to alter incentive mechanisms without otherwise affecting the task.
- Lastly, in *Manuscript 3: "Value-Based Attentional Capture Affects Multi-Alternative Decision Making"*, we aimed to resolve an apparent inconsistency in the literature between two opposing observations with respect to a novel context effect. One of the differences between these findings was the presentation format, with one of them being a perceptually-coded multi-attribute decisions-under-risk task. Using variations of this task and measuring eye movements allowed us to identify the source of this discrepancy and investigate the role of attentional allocation.

2 Context Effects in Decisions from Experience

As discussed in *Models Explaining Context Effects*, one of the core notions that all models designed to explain context effects share is that the attributes are represented clearly by the individuals; Otherwise, they could not perform the calculations required by the models' mechanisms. For example, to determine the strength of the lateral inhibitions of the MDFT model, one needs a numerical value for each of the attributes. This requirement is met in *decisions from description*, one of the most common paradigms used in decision-making research. In that paradigm, participants make decisions on the basis of full descriptions of the choice alternatives and the environment. For example, a decision between lottery A = ($\$4, .8; \$0$) that yields \$4 with probability .8, otherwise nothing, and lottery B = ($\$3, 1$) that yields \$3 for sure.

However, there are certain situations in which, despite the options sharing the very same structure as in decisions from description, the accessibility of the attributes to the participants is limited. For example, individuals could observe samples from the options' outcome distributions, thus inferring their properties. Such an experimental paradigm is called *decisions from experience* and is ideally suited to test the assumption that the clear attribute representation is a necessary condition for context effects to arise. For example, a decision maker might observe the following sequence of seven draws from A: [\$4, \$4, \$4, \$4, \$4, \$0, \$4], and a sequence of eight draws from B: [\$3, \$3, \$3, \$3, \$3, \$3, \$3, \$3]. She would base her decision on these observed outcomes.

Decisions from experience have been of interest to researchers for decades (Edwards, 1961, 1962). However, the development of standardized experiments (e.g., the Iowa gambling task; Bechara, Damasio, Damasio, & Anderson, 1994), advancements in computational modeling of such decisions (Sutton & Barto, 1998), and the discovery of a close link between a model mechanism and neural activity (Schultz, Dayan, & Montague, 1997) increased their popularity substantially. One widely used manifestation of the decisions-from-experience paradigm is the *repeated-choice paradigm*. In the repeated-choice paradigm, individuals repeatedly choose between multiple options. The selected option yields one outcome from that option's outcome distribution which the decision maker obtains, and a new trial begins. In some cases, individuals also obtain the outcome feedback of the non-chosen options (full feedback; see Figure 2.1, for an illustration of the repeated-choice paradigm with full feedback).

Axiomatic economic decision theories do not specify how information upon

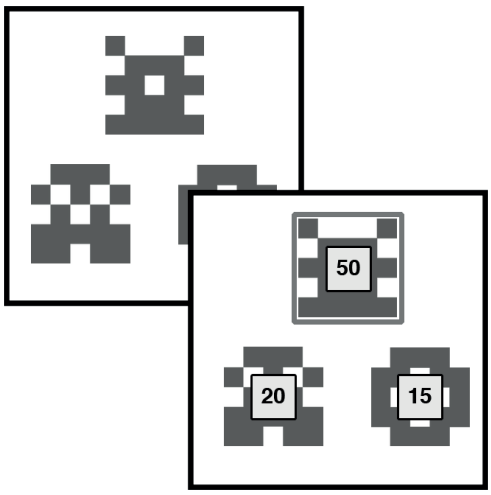


FIGURE 2.1: Illustration of the repeated-choice paradigm with full feedback.

which decisions are based is obtained. In their seminal work, Hertwig et al. (2004) investigated whether the source of that information matters. They used decisions such as between options A and B and contrasted behavior in decisions from description and those from experience. By doing so, they discovered the so-called *decription–experience gap in risky choice* (Hertwig & Erev, 2009). This "gap" is the observation that the typical decisions-from-description finding of as-if overweighting of rare events (e.g., Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) reverses in decisions from experience (i.e., people act as if they underweight rare events). Subsequent research identified several driving forces behind the gap, such as sampling bias (e.g., Fox & Hadar, 2006; Hau, Pleskac, Kiefer, & Hertwig, 2008) or the special role of lotteries with certain outcomes (Glöckner, Hilbig, Henninger, & Fiedler, 2016). Nevertheless, Wulff, Mergenthaler-Canseco, and Hertwig (2017) conclude in their meta-analysis that the gap persists after controlling for the discussed explanatory variables. Consequently, the gap hints towards the possibility that, despite being objectively identical, options' evaluations might be fundamentally different between decisions from description and those from experience.

2.1 Manuscript 1: "How Similarity Between Choice Options Affects Decisions From Experience: The Accentuation of Differences Model"

Spektor, M. S., Gluth, S., Fontanesi, L., & Rieskamp, J. (2017). How similarity between choice options affects decisions from experience: The accentuation of differences model. Manuscript under revision.

If decisions from experience differ in their elicited risk attitudes compared to decisions from description, could the same be true for context effects? In the first manuscript (Spektor, Gluth, et al., 2017), we tried to answer exactly this question. We found that individuals exhibited context effects in decisions from experience. However, these context effects differed from those using decisions from description typically reported in the literature. We propose a new model based on a mechanism that predicts increased attention to salient outcomes, and show that such a model can explain multi-alternative choices in decisions from experience better than simpler learning models.

Our journey started with reviewing how existing cognitive models explain context effects and whether the same mechanisms could be applied to an experience-based setting. As already mentioned above, one of the commonalities amongst all models is the requirement of a clear multi-dimensional representation of options' attributes: Without that property, attributes cannot interact with each other (e.g., Roe et al., 2001; Usher & McClelland, 2004), accessibility cannot be determined (Bhatia, 2013), and attention weights cannot be obtained (Trueblood et al., 2014). Therefore, in order to apply the mechanisms existing in the literature, one would have to assume some mapping from the observations to a numerical multi-attribute representation. A stream of outcomes such as [\$1, \$1, \$0, \$1, \$0, \$0, \$1, \$0] would have to be transformed into the description-equivalent version of equiprobably obtaining \$1 or nothing. In simple cases like in that example, individuals might indeed form such a representation of the option's outcome distribution and use similar cognitive mechanisms as in decisions from description.

However, the plausibility of this assumption vanishes as soon as the outcome distribution becomes more complex. How would a representation in the multi-dimensional attribute space look like if the outcome distribution depends on previous choices (e.g., Biele, Erev, & Ert, 2009), changes constantly (e.g., Boorman, Behrens, & Rushworth, 2011; Daw, O'Doherty, Dayan, Dolan, & Seymour, 2006; Knox, Otto, Stone, & Love, 2012; Nassar, Wilson, Heasly, & Gold, 2010; Speekenbrink & Konstantinidis, 2015), or in which the environment is inherently stochastic and the outcomes are noisy? If we add noise stemming from a half-normal distribution with $\sigma = 0.2$ to the previous paragraph's outcome sequence, we might observe [\$1.31, \$1.38, \$0.03, \$1.46, \$0.21, \$0.57, \$1.14, \$0.08]. It is highly unlikely that individuals represent that sequence as the lottery $(\$1, .5; 0) + \$\mathrm{H}\mathcal{N}(0, 0.2)$. Instead, individuals might as well keep track of how much they like that option and update it with incoming feedback.

A popular way to formalize this intuition is to use *reinforcement-learning models* (Sutton & Barto, 1998) with *temporal-difference learning*. These models keep track of a subjective reward expectation and update it using a fraction of the so-called *reward-prediction error* (i.e., the difference between the reward obtained and the expectation beforehand). Similar to axiomatic economic theories of decision making, reinforcement-learning models assume that both the updating of expectations and the subsequent choice based upon them is done in isolation (i.e., conform to the independence axiom). Whereas some reinforcement-learning models successfully adapted mechanism of decisions-from-description decision-making models, such as risk aversion (e.g., Niv, Edlund, Dayan, & O'Doherty, 2012; Yechiam & Busemeyer, 2005, 2008) or loss aversion (e.g., Steingroever, Wetzels, & Wagenmakers, 2013, 2014; Yechiam & Busemeyer, 2005), so far none of them include a mechanism that could potentially explain context effects. Despite the existence of psychological models that are able to explain specific effects that occur in decisions from experience (Erev & Haruvy, 2005; Erev & Roth, 2014; or Plonsky & Erev, 2017; Plonsky, Teodorescu, & Erev, 2015), they do not address the possibility of context effects either.

In the first manuscript, we proposed a new approach to trial-and-error learning and decision making: We hypothesized that options are not updated and evaluated independently of each other, and expected therefore context effects to arise in decisions from experience as well. In contrast to the existing reinforcement-learning models, we assumed that options interact with each other during the learning process. We proposed a *similarity mechanism* during value updating, according to which the similarity of different options' outcomes inhibits their evaluations, thus making options similar to each other less attractive. Along the lines of the memory bias for salient events (e.g., Ludvig, Madan, & Spetch, 2014; Madan, Ludvig, & Spetch, 2014), the similarity mechanism can be interpreted as an attentional bias for salient outcomes. Saliency captures attention (e.g., Anderson, Laurent, & Yantis, 2011a), and mere attention to options makes them more attractive (Krajbich, Armel, & Rangel, 2010; Lim, O'Doherty, & Rangel, 2011). We formalized this similarity mechanism in the *accentuation-of-differences model* (for formal specifications, see Appendix A) that is otherwise built on a simple reinforcement-learning model.

In a series of three experiments with full feedback (i.e., individuals also obtained the information about forgone outcomes), we designed our stimuli similarly to how typical decision-making-under-risk studies showing context effects designed theirs (e.g., Herne, 1999; Wedell, 1991). This means, we used two core options out of which one had a high probability of yielding a low reward (otherwise nothing) and the other had a low probability of yielding a high reward (otherwise nothing). In one of two choice sets, we added a decoy that, according to traditional context-effect results, should make the safer option appear more attractive, and in the other choice set we added a decoy that should make the riskier option appear more attractive. By contrasting the target choices relative to the competitor choices across these two choice sets, it is possible to quantify the degree to which context matters while controlling for initial preferences. One prediction of the accentuation-of-differences model is that, according to the

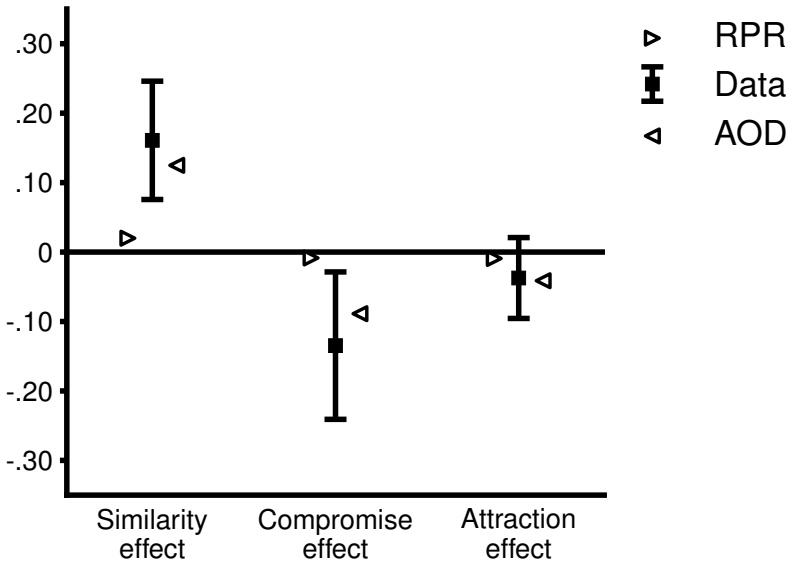


FIGURE 2.2: Empirical context-effect strengths and model predictions for Experiment 1 to 3 (Spektor, Gluth, Fontanesi, & Rieskamp, 2017). Depicted are the differences in relative choice proportions between the target and competitor. Scale ranges from -1 (competitor always chosen) to 1 (target always chosen). Zero indicates no context effect. Experiments are named after the context effect from decisions from description that they resemble. Arrows indicate each of the models' mean context-effect strength across all posterior-predictive simulations. RPR = risk-preference model (runner-up model in all experiments). AOD = accentuation-of-differences model (our proposed model; winner in quantitative model comparison in all experiments). Error bars indicate 95% CI.

similarity mechanism, neither the compromise nor the attraction effect should arise. More important, the model predicts a *reversal* of both effects. The similarity effect, on the other hand, should be present as known from decisions from description.

The main findings across the three experiments are that individuals indeed violated independence in decisions from experience and that they did not obtain an explicit two-dimensional representation of the outcome distributions. In the experiment in which the stimuli were arranged to test the similarity effect, we observed a systematic similarity effect, thus violating independence. In the experiment designed to test the compromise effect, we observed the *reversed* compromise effect. In the experiment designed to test the attraction effect, we observed a weak effect with a tendency in the direction of the *reversed* attraction effect (see Figure 2.2). After each of the experiments, we assessed explicit knowledge about the options' outcome distributions. We found that the responses in these post-questionnaires did not correlate with choices across participants.

In each of the experiments, we rigorously compared the accentuation-of-differences model against three other reinforcement-learning models. We used hierarchical-Bayesian parameter estimation to obtain the posterior distributions and calculate the *widely applicable information criterion* (Watanabe, 2010), quantifying the complexity-penalized model fits relative to each other. Additionally, we tested whether models were inherently able to produce the qualitative choice patterns observed using posterior-predictive simulations (see Steingroever et al., 2014, for a similar approach). In terms of quantitative model fit, our model emerged as the winner across all experiments. It was also best in both qualitatively predicting the observed context effects (see Figure 2.2) and capturing the dynamic development of choices in the tasks.

After demonstrating our model's ability to correctly predict the behavioral phenomena in the experiments designed to test for context effects, we also aimed to assess its generalizability in a standard learning task that was not explicitly designed to elicit context effects. We chose the *Iowa gambling task* (Bechara et al., 1994), as one of the most widely used tasks to study individuals' learning, risk-taking, and impulsivity behavior (cf. Bechara, Damasio, Tranel, & Damasio, 1997; Busemeyer & Stout, 2002). We used published data that compared a control population of individuals with a population of drug abusers (Yechiam, Stout, Busemeyer, Rock, & Finn, 2005). From these data, we used the control group. The challenge for the accentuation-of-differences model was whether it was also able to accurately describe the learning process in such a qualitatively different learning task. As in the previous experiments, our model far outperformed the other candidate models. Parameter estimates across experiments suggest that options yielding distinct outcomes on average attracted more attention and were hence perceived as more attractive than the other options, providing support for the similarity mechanism.

To the best of our knowledge, this study is the first to demonstrate systematic violations of the independence axiom in decisions from experience (but see Tsetsos, Chater, & Usher, 2012, for a perceptual experience-based task), highlighting that explicit attribute-representations are not necessary for context effects to arise. Our work illustrates the shortcomings of standard context-insensitive reinforcement-learning models and incorporates psychological insights into the field of reinforcement learning.

3 Context Effects in Perceptual Decision Making

By definition, preferential choices cannot be classified as "correct" or "incorrect". Therefore, as briefly alluded to in *Context Effects of Preferential Choice*, axiomatic economic theories postulate a set of consistency principles that serve as benchmarks of "rational" human behavior. Although it is possible to redefine choices in terms of accuracy based on the expected values (i.e., whether an individual chose the option with the higher expected value), this accuracy definition would ignore well-known factors that influence behavior, such as risk preferences, loss aversion, or non-linear probability weighting (see Kahneman & Tversky, 1979; Tversky & Kahneman, 1992).

Even though it is not a necessary requirement, indifference between the core options is a highly desirable property in order to demonstrate violations of consistency principles. For example, imagine that option A in the binary choice set $\{A, B\}$ is chosen in 95% of the times. Showing that the relative choice proportion between A and B changes when a new option is added to the choice set becomes almost impossible, as choice proportions are already close to the ceiling/floor. Therefore, researchers have invested substantial efforts to ensure that individuals are indifferent between the core options (e.g., Berkowitsch et al., 2014; Liew et al., 2016). In perceptual tasks, on the other hand, there always is an objectively "correct" answer known to the researcher. If the task is to indicate the longer out of three lines, then every answer can be classified as correct or incorrect by the researcher, independently of the personal preferences of the decision maker. This control over the experimental design allows researchers to create "indifference" much easier by simply increasing the task difficulty.

The use of perceptual tasks as proxy measures for preferential choices is therefore a very appealing perspective. And indeed, their use to study violations of consistency principles has gained popularity in recent years (e.g., Farmer et al., 2017; Trueblood et al., 2013; Trueblood & Pettibone, 2017; Tsetsos et al., 2012; Tsetsos et al., 2016). In various perceptual tasks, context effects known from the preferential-choice literature could be replicated. The attraction effect has been demonstrated most often (e.g., Choplin & Hummel, 2005; Farmer et al., 2017; Parrish et al., 2015; Trueblood et al., 2015; Trueblood et al., 2013), but also the similarity effect has been shown to occur in perceptual tasks (Trueblood et al., 2015; Trueblood et al., 2013). Consequently, it has been suggested that context effects are fundamental decision-making principles not restricted to the consumer-choice domain (Trueblood et al., 2013).

3.1 Manuscript 2: "When the Good Looks Bad: An Experimental Exploration of the Repulsion Effect"

Spektor, M. S., Kellen, D., & Hotaling, J. M. (2017). When the good looks bad: An experimental exploration of the repulsion effect. Manuscript under revision.

What are the boundary conditions of the attraction effect? Is it possible to demonstrate the reversed attraction effect systematically? In the second manuscript (Spektor, Kellen, & Hotaling, 2017), we exploited the advantages of perceptual decision-making tasks to provide an experimental exploration of the *repulsion effect*. We found that the seemingly irrelevant factor of on-screen arrangement of options contributed most to how individuals chose in the perceptual task we used. However, the design of the stimuli also played a non-negligible role: Choice difficulty and distances between the target and decoy in the attribute space moderated whether attraction-, null-, or repulsion effects occured.

As briefly discussed in *Boundary Conditions of Context Effects*, there are some inconsistencies in the literature with respect to the attraction effect: In some cases, it replicates reliably, in other cases researchers observed its reversal, the repulsion effect (Frederick et al., 2014). Despite existing thought experiments about the repulsion effect from almost 30 years ago (Kreps, 1990, p. 28), systematic demonstrations of its occurrence have not been successful so far (Simonson, 2014). The working hypothesis behind repulsion effects is the *tainting hypothesis* (Simonson, 2014, p. 518), according to which similar, clearly inferior alternatives "taint" the attribute space they are located in, thus making the asymmetrically dominating option look less attractive. Intriguingly, Malkoc et al. (2013) showed that unattractive choice sets mitigate the attraction effect, providing support for the general notion that choice-set attractiveness might be the key to the inconsistencies reported in the literature.

What if the attraction effect is susceptible to framing effects? Beyond its mitigation in unattractive choice sets, it is noteworthy that most studies investigating the attraction effect have been conducted using either positive incentives for the participants (e.g., Herne, 1999) or with no incentives at all (e.g., Trueblood et al., 2013). When people make decisions that involve the possibility of a loss, they tend to weight the loss disproportionally compared to gains of equal magnitude (Kahneman & Tversky, 1979). We expected that, much along the lines of the famous "Asian disease problem" (Tversky & Kahneman, 1981), it might be possible to demonstrate a repulsion effect when a decision is framed as a threat of a loss (tainting of the attribute space) and an attraction effect when it is framed as the opportunity of a gain (no tainting of the attribute space). Therefore, we decided to use the rectangle-size task as introduced by Trueblood et al. (2013) to assess this possibility.

A second possible determinant of the repulsion effect might be the specific placement of the decoy in the attribute space. It has been pointed out recently that the *multiattribute linear ballistic accumulator model* (Trueblood et al., 2014) predicts repulsion effects for very close *frequency decoys* (Tsetsos et al., 2015, p. 843). Frequency decoys are decoys that are inferior on the target's stronger attribute, but equally desirable on the target's weaker attribute. Decoys inferior on the target's weaker attribute are referred to as *range decoys*, and decoys weaker on both attributes are called *range-frequency decoys*.[4]

In the first experiment (out of four) we aimed to test the framing-effect intuition of the repulsion effect and explore the influence of the distance between the target and decoy in the attribute space on the context effect observed. Individuals were always presented with three rectangles and were asked to choose the largest of them. In contrast to previous uses of the rectangle-size task (except for Farmer et al., 2017), we made sure that there always was an objectively correct choice. In a *gain-framing condition*, participants started with no initial endowment, but could earn points with each correct or intermediate (i.e., second-largest rectangle) choice. If they always chose the correct rectangle, they ended up with CHF 10 at the end of the experiment. In the *loss-framing condition*, the objective incentive structure was identical. However, we endowed participants with CHF 10 at the beginning of the experiment, and each choice that was not correct deduced a small monetary amount. If all choices were wrong, individuals left with CHF 0 in both conditions. On a within-subject level, we created a balanced, full-factorial design in which we could test for the selective influence of the target–decoy distance in the attribute space. Simultaneously, it allowed us to control for the influence of choice difficulty, decoy type, heuristic decision strategies (e.g., "pick the option that has a different orientation than the other ones" or "pick the larger of the rectangles with identical orientations"; both of these strategies would result in poor performance), and other factors.

In contrast to our initial expectations, we observed neither a framing effect nor an attraction effect in the gain-framing condition. However, we observed a strong repulsion effect in both framing conditions. The repulsion effect persisted after controlling for various potential covariates, such as deliberation time (Pettibone, 2012) or choice difficulty (Huber et al., 2014). Regarding the influence of the target–decoy attribute distance, we found that the further the two options were away from each other, the weaker the repulsion effect became. We confirmed these results in two follow-up experiments in which we removed the framing manipulation and altered the stimulus design slightly.

After three experiments of observing only repulsion effects, our experimental results stood at odds with the attraction effect reported in the literature. Experiment 4 aimed at bridging that gap and identifying the moderators that promoted the occurrence of repulsion effects in our experiments and attraction effects in previous usages of the rectangle-size task. We identified three potential factors. These were

[4]This terminology was originally introduced by Huber et al. (1982).

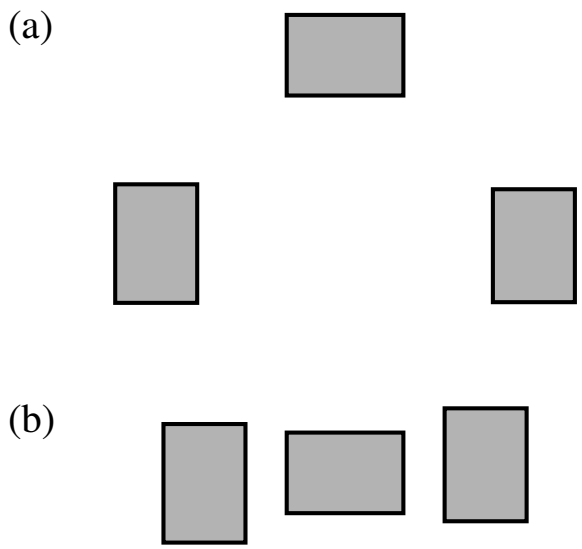


FIGURE 3.1: Schematic illustration of the two stimulus displays we used in Spektor, Kellen, and Hotaling (2017). (a) depicts the arrangement used in Experiment 1 to 3 and 4b and (b) the one used in Experiment 4a. Note that a jitter was applied to both depicted stimulus displays.

- *Stimulus design*; In contrast to our experiments, previous demonstrations always had equally sized core rectangles and the target–decoy attribute distance was not manipulated.
- *Stimulus display*; In our experiments, the rectangles were arranged in a large triangle on-screen with horizontal and vertical jitter. In the original study by Trueblood et al. (2013), the rectangles were arranged close to each other in a horizontal line with a slight vertical jitter (see Figure 3.1 for a schematic illustration of the two arrangements).
- *Absolute rectangle size*: Our rectangles were, on average, around 250 pixels $\times$ 165 pixels large, whereas previous studies used rectangles with average sizes of 80 pixels $\times$ 50 pixels.

The full hypothetical space of potential experimental designs that explores all factor combinations of our experiments and those in the literature thus spans three dimensions with two levels each: stimulus design $\times$ stimulus display $\times$ absolute rectangle size.[5] We hypothesized that the absolute rectangle sizes' influence on decisions was negligible and decided thus to explore the remaining three cells of that experimental-design space.

In Experiment 4a, we used exactly the same stimulus display as in Trueblood et al. (2013) and manipulated between participants whether they completed the original trials that Trueblood et al. (2013) used (direct-replication condition) or a new set of now-smaller rectangles that were generated similarly to our

[5] In principle, these dimensions are continuous. For the sake of simplicity, we treated them discretely.

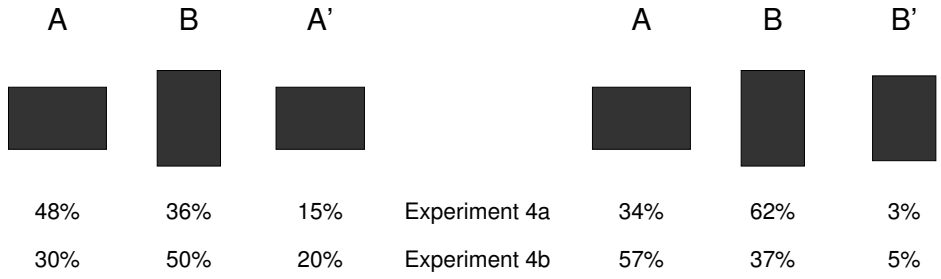


FIGURE 3.2: Illustration of the stimulus-display induced context-effect reversal. Depicted are mean choice proportions of the trials with the IDs 243 and 333 from Spektor, Kellen, and Hotaling (2017, Experiment 4, direct-replication condition). Options A and B are identical in both choice sets. Experiment 4a and 4b differ in the arrangement of stimuli on-screen.

Experiment 1 to 3 (new-trials condition). In Experiment 4b, we collected data in the direct-replication condition with a stimulus display that we used in our previous experiments.

In Experiment 4a's direct-replication condition, we found an attraction effect that disappeared in the new-trials condition. The attribute-distance manipulation seemed to suppress the global context-effect analysis; We found that at close attribute distances, individuals seemed to exhibit repulsion effects and attraction effects at large attribute distances. Strikingly, the very same direct-replication trials arranged on-screen in a large triangle resulted in a strong repulsion effect (see Figure 3.2 for an example pair of choice sets in which the context effect flips completely between Experiment 4a and 4b).

As a last step, we fitted the multiattribute linear ballistic accumulator model (Trueblood et al., 2014) to the data of Experiment 3 and 4. The model was able to account for the observed repulsion effects by placing substantially greater weight on negative comparisons relative to the positive counterparts. However, none of the extant context-effect models are able to provide an a-priori account of why stimulus display should have such a strong effect on individuals' choices.

In sum, we found that both stimulus design and stimulus display contributed to the behavior of decision makers. Small deviations from the exact stimulus design used by Trueblood et al. (2013) resulted in the disappearance of the attraction effect. As soon as the options were arranged differently on-screen, we found highly consistent and robust repulsion effects. We conclude that in the rectangle-size task, researchers are much more likely to observe a repulsion than an attraction effect, as the latter requires the joint occurrence of several specific factor combinations, and the former arises in all other cases.

4 Context Effects in Decisions Under Risk

The first computational model to simultaneously account for all three context effects, MDFT (Roe et al., 2001), explained the attraction effect using lateral inhibition. Although lateral inhibition was in principle neurally motivated (e.g., lateral inhibition seems to be the neural mechanism behind optical illusions; see Eagleman, 2001), Roe et al. (2001) allowed activations to go below 0: Consistently negative evaluations of the decoy relative to the target lead to decoy activations below 0 and thus boost the target's input through the resulting disinhibition. This mechanism was soon criticized due to its neural implausibility (Usher & McClelland, 2004, p. 762), as firing rates of neurons have an intrinsic lower boundary of 0 s^{-2} (i.e., they do not fire at all) and hence cannot have a negative activation. Incidentally, the study of violations of economic consistency principles recently gained prominence in the field of neuroscience as well (e.g., Chau, Kolling, Hunt, Walton, & Rushworth, 2014; Gluth, Hotaling, & Rieskamp, 2017; Hunt, Dolan, & Behrens, 2014; Klein, Ullsperger, & Jocham, 2017; Louie, Grattan, & Glimcher, 2011; Mohr et al., 2017; Palminteri, Khamassi, Joffily, & Coricelli, 2015). The corresponding models that have been proposed are designed with biological plausibility in mind. Among the many theories, *divisive normalization* (Louie, Khaw, & Glimcher, 2013) emerged as a particularly successful framework of context-dependent decision making.

Initially, divisive normalization was proposed as a mechanism of sensory perception (see Carandini & Heeger, 2011) and later extended to decision making (Louie et al., 2013). The basic notion underlying divisive normalization is rather simple: Neural firing rates adapt to the decision context in order to be able to effectively discriminate between the options at hand. For example, if neural firing rates discriminate well between two low values (e.g., \$5 and \$8) and do not adapt to the decision context, then their firing rates might reach the limit with values such as \$500 and \$800 and they would not be able to discriminate between \$5,000 and \$8,000. On the other hand, if they are able to discriminate well between high values, then they would not be able to do so between low values as the firing rates would be very close to the lower boundary. However, empirically it is evident that discrimination is possible in all these value ranges. With divisive normalization, firing rates would scale down proportionally to the sum of all values under consideration. In psychological terms, (full) divisive normalization of values transforms utilities of options onto a relative value scale (e.g., Stewart, Chater, & Brown, 2006; see Vlaev, Chater, Stewart, & Brown, 2011, for an overview over different types of value encoding).

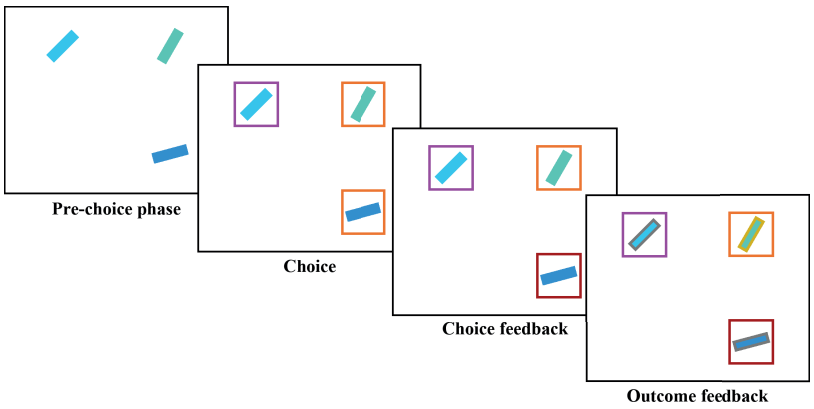


FIGURE 4.1: Example trial of the task originally introduced by Chau, Kolling, Hunt, Walton, and Rushworth (2014) and used in Gluth, Spektor, and Rieskamp (2017). Magenta frames indicate that an option is not available. Golden (gray) frames around rectangles indicate options that yielded (did not yield) their respective reward.

One intriguing property of the model is that it predicts violations of the independence principle *without* a direct involvement of the multi-dimensional attributes of the options: As the *sum* of subjective values in the choice set increases, the relative discriminability between better options decreases. For example, the relative choice accuracy of choosing a higher-valued option A with utility $U(\mathrm{A}) = \$8$ and a lower-valued option B with utility $U(\mathrm{B}) = \$7.50$ will be lower in a choice set consisting of $\{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$ than in a choice set comprising $\{\mathrm{A}, \mathrm{B}, \mathrm{D}\}$ whenever $U(\mathrm{C}) > U(\mathrm{D})$. Such a context effect would challenge not only axiomatic economic theories, but also psychologically motivated models that assume some attribute-comparison process behind the occurrence of context effects.

Using a three-alternative food-snack-choice task, Louie et al. (2013) reported that distractor-option value had a negative impact on relative choice accuracy: Higher-valued distractors impaired relative choice accuracy between the high-value and low-value core options more strongly than lower-valued distractors.[6] Shortly after this observation, Chau et al. (2014) reported the diametrically opposite pattern: Compared to lower-valued distractors, higher-valued distractors *improved* relative choice accuracy between the other two options. Chau et al. (2014) proposed a biophysical model (a model quite similar to the leaky, competing accumulator model; Usher & McClelland, 2001) to explain their results. With the observations of Louie et al. (2013) and Chau et al. (2014) at odds, it was unclear whether value is positively or negatively related to relative choice accuracy.

[6]It should be noted that, although it is not a necessary requirement, Louie et al.'s (2013) distractors always had the lowest value of the three options. In other words, even the highest-value distractor had a lower value than the lowest-value core option.

4.1 Manuscript 3: "Value-Based Attentional Capture Affects Multi-Alternative Decision Making"

Gluth, S., Spektor, M. S., & Rieskamp, J. (2017). Value-based attentional capture affects multi-alternative decision making. Manuscript under revision.

In contrast to Louie et al.'s (2013) experimental task, Chau et al. (2014) used a paradigm in which options' attributes were presented explicitly and had to be integrated by the participants to make decisions. Could it be that the results reported by Chau et al. (2014) stem from the multi-attribute nature of the task? This question was the starting point for the third manuscript (Gluth, Spektor, & Rieskamp, 2017). In the end, we found no systematic influence of distractor value on relative choice accuracy, a result incompatible with both divisive normalization and the original findings of Chau et al. (2014). Instead, we observed behavior in line with a simple attentional mechanism.

The experimental paradigm introduced by Chau et al. (2014) is an interesting mix of a perceptual, decisions-from-description, and decisions-from-experience task. Essentially, it is a decisions-under-risk task in which participants choose between three two-outcome lotteries.[7] These lotteries yield some amount between £2 and £12 with probabilities ranging from $\frac{1}{8}$ to $\frac{7}{8}$, or otherwise nothing. However, the options are not presented numerically to the participants but instead they are represented by colored, rotated rectangles whose colors (on a green–blue scale) code the outcome magnitudes and whose angles (on a horizontal–vertical scale) code the outcome probabilities (see Figure 4.1 for an example trial of the task). In contrast to the vast majority of decisions-under-risk tasks used in the literature (e.g., Abdellaoui, Bleichrodt, & Paraschiv, 2007; Birnbaum, 2008), individuals obtain the outcome feedbacks of all three options on each trial. Each trial lasts up to 1,600 ms in which individuals can make a choice. The important manipulation in Chau et al.'s (2014) study was that after the first 100 ms, one of the options was declared unavailable (distractor), independently of its value. Chau et al. (2014) observed that higher-valued distractors *improved* the relative choice accuracy between the higher-valued remaining option and the lower-valued remaining option.

With such an experimental set-up, the options can be represented in a two-dimensional attribute space (similar to Figure 1.1). We hypothesized that this multi-attribute nature of the task might have led to the observed accuracy-improving effect of distractor value on choice accuracy. For example, comparison processes between the different orientations and colors can lead to attraction-like effects or phantom-decoy effects.[8] To test the hypothesis that the apparent contradiction of Louie et al.'s (2013) and Chau et al.'s (2014) results stems from the different presentation formats, we used Chau et al.'s (2014) task to

[7] Half of the trials in the study involved only two two-outcome lotteries.

[8] The effect would only be "attraction-like" because the third option is not available for choice.

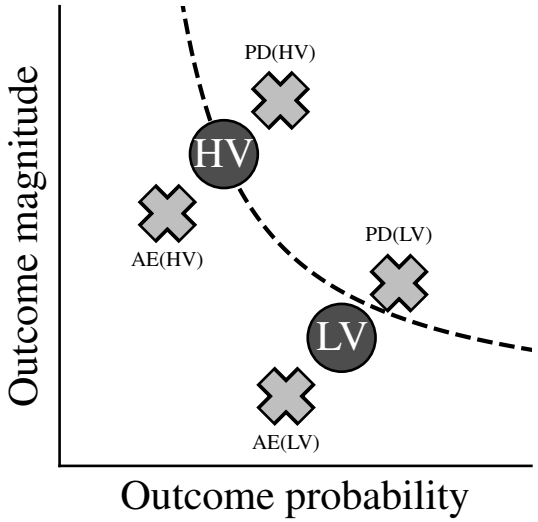


FIGURE 4.2: Arrangement of the "special trials" (Gluth, Spektor, & Rieskamp, 2017, Experiment 1–3). AE(HV) = attraction-like effect placement of the distractor (i.e., inferior to the higher-valued option [HV]). PD(HV) = phantom-decoy placement of the distractor (i.e., superior to HV). The same terminology applies to the lower-valued option (LV). Dashed line indicates the curve along all options with the same expected value as HV lie.

systematically test for these attribute-dependent context effects. We identified all possible combinations of core options for which it was possible to add an asymmetrically dominating or dominated distractor to either of the options. Using these so-called *special trials* (see Figure 4.2 for an illustration of the special-trials arrangement in the two-dimensional attribute space), we could discriminate between various theoretical and empirical accounts of the data as they predict different choice patterns (see Figure 4.3 for qualitative predictions of each effect/model-based account and the aggregated data).

In a first experiment, participants completed Chau et al.'s (2014) original set of trials with the novel special trials interleaved. To our surprise, we neither observed the original effect observed by Chau et al. (2014) nor behavior consistent with divisive normalization. *Relative* choice accuracy was not systematically affected by the distrator's value, *absolute* choice accuracy, on the other hand, was. In contrast to divisive normalization and Chau et al.'s (2014) biophysical model, *value-based attentional capture* (Anderson, Laurent, & Yantis, 2011b) predicts such a pattern in which absolute choice accuracy is affected by the distractor's value, but relative choice accuracy is not (see Figure 4.3). Value-based attentional capture is the notion that value-laden irrelevant stimuli attract attention and impair goal-directed actions. In other words, options with higher values receive more attention, thereby "stealing" the time needed to attend to the other options and choose accurately. As further evidence in favor of such an account, we found that individuals were slower to respond when distractors had higher values, thus missing the response deadline, and tended to choose the unavailable distractor more often.

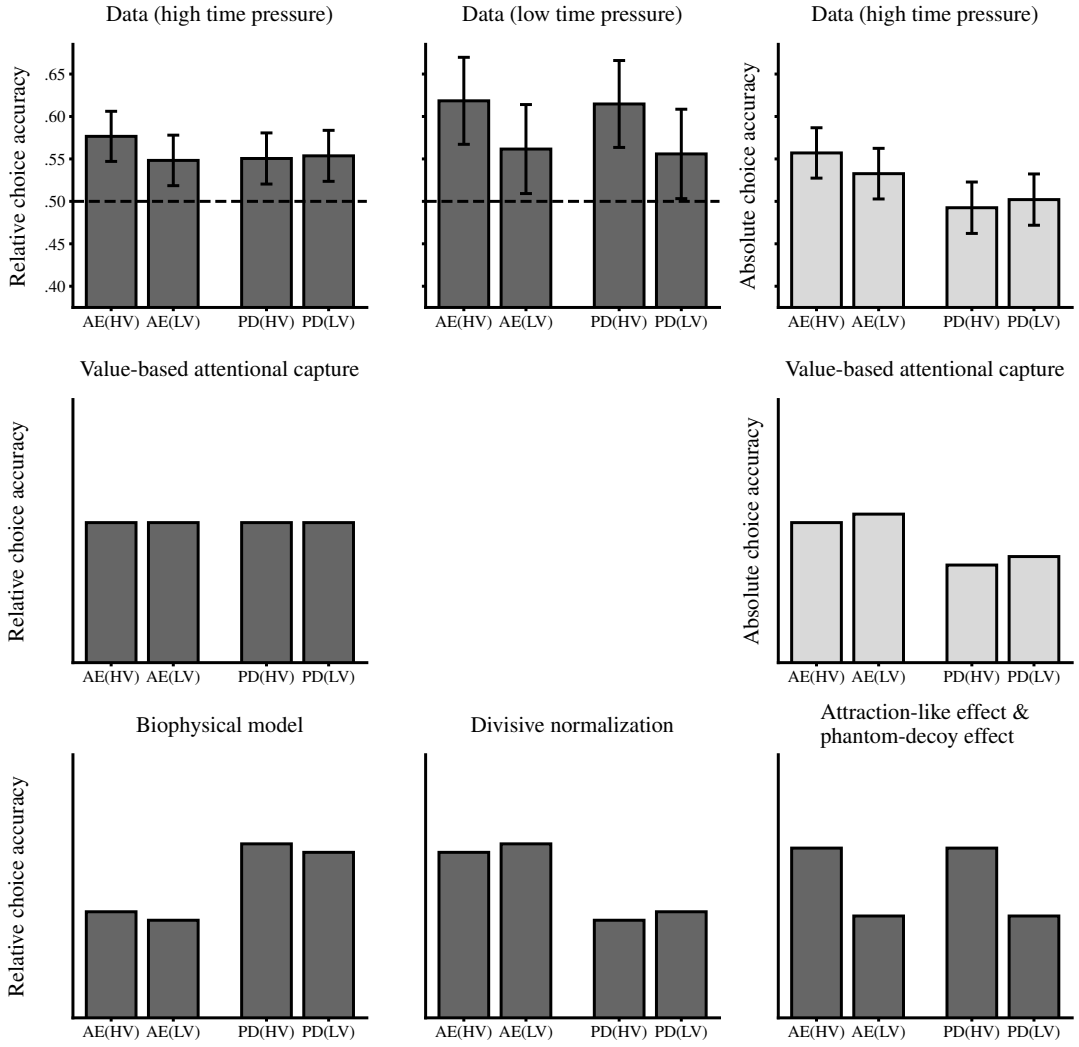


FIGURE 4.3: Observed choice accuracies and predicted qualitative patterns in "special trials" (Gluth, Spektor, & Rieskamp, 2017, aggregated across all experiments in the respective conditions). AE(HV) = attraction-like effect placement of the distractor (i.e., inferior to the higher-valued option [HV]). PD(HV) = phantom-decoy placement of the distractor (i.e., superior to HV). The same terminology applies to the lower-valued option (LV). Dark-gray bars indicate relative choice accuracy. Light-gray bars indicate absolute choice accuracy. Error bars indicate 95% CI.

In two follow-up experiments, we tested the value-based-attentional-capture explanation of the data. We expected the distracting effect of high-valued distractors to disappear as soon as the high time pressure (a total of 1,600 ms per trial) is relieved. In Experiment 2, we manipulated deliberation time between participants to assess its influence. In the high-time-pressure condition (i.e., replication of Experiment 1), we did not observe an influence of distractor value on relative choice accuracy. However, the effects on absolute choice accuracy replicated. Crucially, as soon as individuals had enough time to ignore the attention-capturing effect of the distractor, its influence on choice accuracy disappeared. In the special trials, we observed behavior that is compatible with an attraction-like effect and a phantom-decoy effect (see Figure 4.3). In Experiment 3, we recorded participants' eye movements as a proxy for attention. Otherwise, participants completed the same experiment as in Experiment 1 and in the high-time-pressure condition of Experiment 2. Behaviorally, we found the same pattern as in the previous experiments. More important, we found that individuals looked longer at higher-valued distractors which in turn reduced absolute choice accuracy. Across individuals, we observed that a stronger eye-fixation distraction (i.e., the influence of distractor value on relative fixation duration) correlated with the negative influence of the distractor's value on accuracy. Analyzing the special trials across all experiments with high time pressure provided further support for value-based attentional capture and against alternative explanations.

In sum, we found no effect of distractor value on relative choice accuracy. Instead, we observed a systematic influence of distractor value on absolute choice accuracy, compatible with a value-based-attentional-capture account of the data. An experimental reduction of time pressure and the analysis of eye-movement patterns supported this interpretation of the results.

5 Discussion

Decades of research have shown that context has a non-negligible influence on decisions, thus challenging axiomatic economic decision-making theories. However, with the discovery of many boundary conditions of context effects, their robustness has recently been cast into question. Presentation format has been identified as one of these boundary conditions that can make context effects disappear or even reverse. Yet, a systematic exploration of the influence of presentation format on context effects has not been conducted so far and the cognitive processes underlying such a behavioral change remain elusive. In the present dissertation, I therefore aimed to extend our understanding of the dependency of choices on context and presentation format. In the first manuscript, we transferred typical description-based context-effect constellations to a decisions-from-experience paradigm. We proposed an attention-based similarity mechanism, as formalized in our accentuation-of-differences model, that predicts violations of the independence axiom without the need for obtaining a clear multi-dimensional representation of options' attributes. We demonstrated that individuals violated independence in line with our model's predictions and that they indeed did not obtain explicit knowledge about the options' properties. In the second manuscript, we rigorously evaluated the boundary conditions of the attraction effect in a perceptual decision-making task. We identified that arrangement of stimuli on-screen had an influence on choices that far surpassed that of the stimulus design. Only under a very specific combination of stimulus arrangement and design could we obtain the attraction effect. In all other cases, repulsion effects—reversed attraction effects—were ubiquitous. Finally, in the third manuscript, we found that novel "attribute-free" context effects do not occur in a decisions-under-risk task with perceptually coded attributes. When individuals were put under time pressure, traditional context effects did not arise either. Value-based attentional capture provided a coherent explanation for all behavioral results and was further supported by eye-movement patterns. In sum, context has a prevailing influence on human decision making. Yet, the manifestation of this context dependency is itself context dependent, and different presentation formats pose boundary conditions for their occurrence.

5.1 The Special Role of Attention

In this part of my dissertation, I want to briefly discuss the role of the one psychological process that seems to be especially important across all presented manuscripts: *attention*. Attention is one of the most-widely investigated topics in psychology. Phenomena related to attention, such as the *Stroop effect* (Stroop, 1935), have been known for more than 80 years now. The important role of attention in value formation (e.g., Lim et al., 2011; Shimojo, Simion, Shimojo, & Scheier, 2003) and the role of value in attention allocation (e.g., Anderson et al., 2011b) was, however, discovered only recently. Many formal models rely on attentional mechanisms to explain context effects (e.g., Roe et al., 2001; Usher & McClelland, 2004). These attentional mechanisms represent the computational micro-process of a decision. In contrast to such a formal definition of attention, I want to highlight the importance of attention on a higher level of cognition, such as attentional capacity and which aspects of the stimuli attention is directed towards.

The clearly most direct link to attention has been drawn in manuscript three in which we manipulated attentional capacity experimentally and measured attention using an eye-tracking device. There, value-based attentional capture provided a coherent explanation of the behavioral phenomena we observed: Higher-valued distractor options attracted individuals' attention, making their decisions slower and leading them to choose the distractor, even though it was declared unavailable. When individuals had enough time to direct their attention away from such attractive, yet unavailable options, then their choice accuracy was unaffected by the distractor.

Attention could have been the driving force behind the first manuscript's results as well. From a computational point of view, attention is inherently integrated in the accentuation-of-differences model that we proposed. Its psychological interpretation corresponds to an attentional bias for salient outcomes. Whereas value-based attentional capture is a rather novel discovery, saliency has been known to capture attention for a much longer time (see Anderson et al., 2011a), and merely fixating options makes them appear more attractive (Krajbich et al., 2010; Lim et al., 2011). This psychological interpretation of the similarity mechanism can be tested in a variety of ways. For example, saliency could be experimentally directed away from the option yielding the most distinct outcomes to an average outcome by, for example, increasing its feedback's contrast or displaying the numbers in a larger font or different color. One could also use a psychophysiological measure of attention, such as the eye-movement patterns as we did in the third manuscript, or manipulate attention by using a sequential display of outcomes.

A sequential display of stimuli has very recently been used in a study of the attraction effect in the rectangle-size task (Trueblood & Dasari, 2017). Trueblood and Dasari report a substantial effect of presentation order, with effects ranging from strong attraction to strong repulsion effects. Only after aggregating all presentation orders did the authors observe a very weak attraction

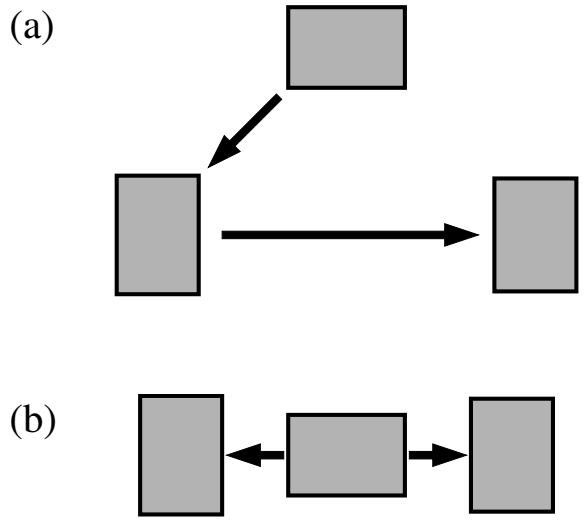


FIGURE 5.1: Hypothetical time-courses of inspection. (a) depicts schematically the arrangement used in Experiment 1 to 3 and 4b (Spektor, Kellen, & Hotaling, 2017) and (b) the one used in Experiment 4a.

effect. While we presented all stimuli simultaneously in the second manuscript, a weaker variant of potential presentation-order effects can nevertheless be explored. Across trials, we completely randomized the positions at which the options were located. Still, in the experiments using the triangular arrangement (Experiment 1–3 and 4b), individuals chose the option on top (39%) more often than the options on the left (31%) and on the right (30%). Similarly, the option in the middle of a horizontal-in-line arrangement (as used in Experiment 4a) was also preferred to the outer options (37%, 31%, and 32% choices for the middle, left, and right option, respectively). Consider the following: Between trials, individuals most likely look at the center of the screen. In case of a triangular arrangement, it is natural to revert to the natural reading order, thus evaluating the rectangles in the order top–left–right, as depicted in Figure 5.1 (a). In case of a horizontal-in-line arrangement, it is easiest to simply continue fixating on the center of the screen, see Figure 5.1 (b).

The critical question is whether this preference for specific locations had an effect on the occurrence of the context effects. Following the dissection as implemented by Trueblood and Dasari (2017), I separated the context-effect analysis into all six possible "presentation orders" (according to the hypothesized orders depicted in Figure 5.1). By inspecting the results of this dissection, a consistent pattern across all experiments seemed to be whether the target was located on top/in the middle or not. On top of the experiment-to-experiment fluctuations in absolute size of the attraction/repulsion effects, relative choice shares of the target (a measure of context-effect magnitude as introduced by Berkowitsch et al., 2014) were higher if the target was on top/in the middle

than if it was not.[9]

To quantify this observation, I collapsed the six orders back to two levels (target on top/in the center of the screen vs. target in a peripheral position) and analyzed whether targets were chosen more often than competitors, aggregated across all experiments. The benchmark for that analysis is the repulsion-effect statistic disregarding the position of the target: $t(265) = 9.40$, $d_z = 0.58$. And indeed, this analysis revealed that if the target was *not* in the middle position, the repulsion effect became even stronger, $t(265) = 12.86$, $d_z = 0.79$. More interestingly, if the target *was* in the middle position, the repulsion effect disappeared and the sign reversed (i.e., if anything, then the target was chosen more often than the competitor), $t(265) = 1.70$, $d_z = 0.10$. These results are somewhat consistent with those reported by Trueblood and Dasari (2017). In their study, attraction effects arose when the target was presented second and repulsion effects arose in three out of four other presentation orders. An investigation using eye tracking would help to clarify how inspection order and spatial locations of stimuli interact with each other to elicit repulsion or attraction effects (see Noguchi & Stewart, 2014, for a similar approach in the consumer-choice domain).

5.2 Theoretical Challenges

In the second manuscript, we identified that target–decoy attribute distance influenced choices in every experiment. Interestingly, the effect was consistent across both stimulus displays and the three stimulus-design manipulations we investigated: Whenever the decoy moved further away from the target in the attribute space, the observed context effect moved in the direction of the attraction effect; A weak repulsion effect disappeared and reversed (Experiment 4a, new-trials condition), and a strong repulsion effect became weaker (Experiment 1–3). Even though we have not tested the new-trials condition in a stimulus-display format as in the other experiments, the results from Experiment 4b suggest that we would observe the identical pattern. So far, none of the available cognitive models predict such an effect of target–decoy attribute distance, as most of them predict a monotonic or step-wise effect (Bhatia, 2013; Tsetsos et al., 2010; but see Trueblood et al., 2014, for a model that predicts a curvilinear relationship). Yet, whether the same distance-dependency also holds for preferential choices remains to be seen.

Due to lack of empirical support, the prediction of repulsion effects has been perceived as rather detrimental for cognitive models (see Tsetsos et al., 2015). Trueblood et al. (2015) argued that the specific placement of options discussed by Tsetsos et al. (2015) would lead to a similarity-like effect (i.e., individuals confuse the decoy and the target) and that this effect would arise in the rectangle-size task. Our results of the second manuscript provide evidence against both

[9]Note that due to full randomization of option placement, the analysis of the relative choice share of the target could be biased since it cannot be guaranteed that the matched choice sets are included equally often within each presentation order and individual.

notions. First, we observed universal repulsion effects independently of decoy type and distance. Second, we can refute the similarity-like-effect interpretation due to the fact that even in the most difficult choice sets, individuals were able to clearly discriminate the target from the decoy and chose the former more often than the latter (see Spektor, Kellen, & Hotaling, 2017, Supplemental Material). To date, there are no formal accounts of the repulsion effect available. More important, a general theory for both perceptual and preferential decision-making tasks has to take the influence of stimulus display into account.

In the first manuscript, we proposed the accentuation-of-differences model as a process model of learning and decision making by trial-and-error. We have shown that it outperforms traditional reinforcement-learning models and generalizes to the Iowa gambling task with full feedback. Beyond the model's ability to qualitatively predict and quantitatively fit the observed behavioral patterns, it makes a very distinct prediction with respect to the co-occurrence of outcomes. If outcomes are not drawn independently from their respective options' distributions but are tied to some underlying "event" (similar to the illustration in Busemeyer & Townsend, 1993, Figure 5), then our model predicts that options whose outcomes are tied to anti-correlated events will be perceived as more attractive. A second critical test for our model is the assumption that context influences the perceptions of value during learning, but that decisions themselves are then based on an independent evaluation of the resulting expectations. This prediction is most likely difficult to test only behaviorally, but psychophysiological or neural activity during learning might shed light on the empirical content of this assumption.

5.3 Future Empirical Directions

In recent years, increasingly many moderators and boundary conditions of context effects have been identified. However, the empirical evidence for some of these moderators is inconsistent. For example, compared to preferential tasks, it was reported that the attraction effect is attenuated in perceptual tasks (e.g., Farmer et al., 2017; Trueblood et al., 2013). In the second manuscript, we observed that the attraction effect is mostly not only attenuated, but even not present in the very same perceptual task; Instead, we found the repulsion effect to occur most of the time. Stimulus display (i.e., the arrangement of stimuli on-screen) had the by-far strongest effect on how context influences decisions. As proposed in *The Special Role of Attention*, attention might be the crucial cognitive process underlying this dramatic change of behavior. Future research should clarify whether this is the case. For example, options could be presented sequentially and in different locations of the screen, or attention could be directed towards a specific option by showing a salient fixation cross at the location at which the option of interest will appear.

In the third manuscript, we observed an attraction-like effect and the phantom-decoy effect (see Figure 4.3) when individuals were given enough deliberation time, but a null effect when they were put under time pressure. This result

is in line with previous observations (Pettibone, 2012; Trueblood et al., 2015). However, Trueblood and Pettibone (2017) reported a complete reversal of the phantom-decoy effect using a purely perceptual task. Even though our task was in essence a decisions-under-risk task, the fact that we used rectangles to code the lotteries' properties makes it rather similar to the rectangle-size task. So far, it remains unclear which boundary conditions lead to such a reversal and at which point this reversal takes place. One approach to investigate this would be a between-subject manipulation of whether the task is framed as preferential or perceptual, for instance, as "pick the greenest-and-most-upright rectangle". Such an approach has successfully been used to identify differences between preferential and perceptual tasks (e.g., Dutilh & Rieskamp, 2016; Farmer et al., 2017; Zeigenfuse, Pleskac, & Liu, 2014).

In the first manuscript, we have shown that individuals violate independence in a repeated-choice task with full feedback. We demonstrated that a similarity mechanism as formalized in our newly proposed accentuation-of-differences model provides a good account of the data. Yet, it is unclear how much the results depend on the fact that we provided full feedback. The two repeated-choice paradigms with full and partial feedback are more similar to each other than to other experience-based paradigms or, more dramatically so, decisions from description (Camilleri & Newell, 2011). However, differences between full and partial feedback in model-based characterizations of the learning processes (e.g., Camilleri & Newell, 2011; Rakow, Newell, & Wright, 2015; Yechiam & Busemeyer, 2005, 2006) and in choice phenomena (e.g., Plonsky & Erev, 2017; Plonsky et al., 2015) have been reported as well. The exploration–exploitation dilemma that is effectively eliminated in full-feedback situations plays an important role in partial-feedback situation. I expect the similarity mechanism to gain in importance in such settings, as within-option integrations become increasingly difficult to perform because individuals do not obtain so many observations from the options' outcome distributions. Simply comparing last-seen outcomes, on the other hand, is cognitively less demanding. Therefore, people might increasingly rely on the similarity mechanism in the partial-feedback repeated-choice paradigm.

5.4 Conclusion

Taking everything together, I found that presentation format plays an important role in the occurrence of context effects. Although speculative, the results reported in the three manuscripts can be re-interpreted as shifts in attention: to particularly salient outcomes (Manuscript 1), to option comparisons that determine the ultimate choice (Manuscript 2), or to unavailable-but-attractive distractor options (Manuscript 3). Direct experimental manipulations of attention and saliency or psychophysiological measurements of attention can shed light on such an explanatory framework. If successful, the attentional explanation could de-mystify the inconsistent and seemingly incompatible findings of boundary conditions in context-effect research.

References

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, *53*, 1659–1674. doi:10.1287/mnsc.1070.0711

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011a). Learned value magnifies salience-based attentional capture. *PLoS ONE*, *6*, e27926. doi:10.1371/journal.pone.0027926

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011b). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*, 10367–10371. doi:10.1073/pnas.1104047108

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7–15. doi:10.1016/0010-0277(94)90018-3

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*, 1293–1295. doi:10.1126/science.275.5304.1293

Berkowitsch, N. A. J., Scheibehenne, B., & Rieskamp, J. (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General*, *143*, 1331–1348. doi:10.1037/a0035159

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, *120*, 522–543. doi:10.1037/a0032457

Biele, G., Erev, I., & Ert, E. (2009). Learning, risk attitude and hot stoves in restless bandit problems. *Journal of Mathematical Psychology*, *53*, 155–167. doi:10.1016/j.jmp.2008.05.006

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, *115*, 463–501. doi:10.1037/0033-295X.115.2.463

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS Biology*, *9*, e1001093. doi:10.1371/journal.pbio.1001093

Busemeyer, J. R. & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment*, *14*, 253–262. doi:10.1037/1040-3590.14.3.253

Busemeyer, J. R. & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459. doi:10.1037/0033-295X.100.3.432

Camilleri, A. R. & Newell, B. R. (2011). 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*, 377–384. doi:10.3758/s13423-010-0040-2

Carandini, M. & Heeger, D. J. (2011). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*, 51–62. doi:10.1038/nrn3136

Chau, B. K. H., Kolling, N., Hunt, L. T., Walton, M. E., & Rushworth, M. F. S. (2014). A neural mechanism underlying failure of optimal choice with multiple alternatives. *Nature Neuroscience*, *17*, 463–70. doi:10.1038/nn.3649

Choplin, J. M. & Hummel, J. E. (2005). Comparison-induced decoy effects. *Memory & Cognition*, *33*, 332–343. doi:10.3758/BF03195321

Chung, X. H.-k., Sjo, T., Lee, H.-j., Lu, Y.-t., Tsuo, F.-y., Chen, X.-s., ... Huang, C.-y. (2017). Why do irrelevant alternatives matter? An fMRI-TMS study of context-dependent preferences. *The Journal of Neuroscience*, *37*, 11647–11661. doi:10.1523/JNEUROSCI.2307-16.2017

Daw, N. D., O'Doherty, J. P., Dayan, P., Dolan, R. J., & Seymour, B. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876–9. doi:10.1038/nature04766

Debreu, G. (1960). R. D. Luce, Individual Choice Behavior: A Theoretical Analysis. *American Economic Review*, *50*, 186–188.

Dhar, R. & Simonson, I. (2003). The effect of forced choice on choice. *Journal of Marketing Research*, *1*, 303–304. Retrieved from http://www.jstor.org/stable/30038845

Doyle, J. R., O'Connor, D. J., Reynolds, G. M., & Bottomley, P. A. (1999). The robustness of the asymmetrically dominated effect: Buying frames, phantom alternatives, and in-store purchases. *Psychology and Marketing*, *16*, 225–243. doi:10.1002/(SICI)1520-6793(199905)16:3<225::AID-MAR3>3.0.CO;2-X

Dutilh, G. & Rieskamp, J. (2016). Comparing perceptual and preferential decision making. *Psychonomic Bulletin & Review*, *23*, 723–737. doi:10.3758/s13423-015-0941-1

Eagleman, D. M. (2001). Visual illusions and neurobiology. *Nature Reviews Neuroscience*, *2*, 920–926. doi:10.1038/35104092

Edwards, W. (1961). Probability learning in 1000 trials. *Journal of Experimental Psychology*, *62*, 385–394. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13889318

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors*, *4*, 59–73.

Erev, I., Ert, E., Plonsky, O., Cohen, D., & Cohen, O. (2017). From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychological Review*, *124*, 369–409. doi:10.1037/rev0000062

Erev, I. & Haruvy, E. (2005). Generality, repetition, and the role of descriptive learning models. *Journal of Mathematical Psychology*, *49*, 357–371. doi:10.1016/j.jmp.2005.06.009

Erev, I. & Roth, A. E. (2014). Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences*, *111*, 10818–10825. doi:10.1073/pnas.1402846111

Evangelidis, I., Levav, J., & Simonson, I. (2017). The asymmetric impact of context on advantaged versus disadvantaged options. *Journal of Marketing Research*. Advance online publication. doi:10.1509/jmr.14.0483

Farmer, G. D., Warren, P. A., El-Deredy, W., & Howes, A. (2017). The effect of expected value on attraction effect preference reversals. *Journal of Behavioral Decision Making*, *30*, 785–793. doi:10.1002/bdm.2001

Fox, C. R. & Hadar, L. (2006). "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, *1*, 159–161.

Frederick, S., Lee, L., & Baskin, E. (2014). The limits of attraction. *Journal of Marketing Research*, *51*, 487–507. doi:10.1509/jmr.12.0061

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*, 486–508. doi:10.1037/a0040103

Gluth, S., Hotaling, J. M., & Rieskamp, J. (2017). The attraction effect modulates reward prediction errors and intertemporal choices. *Journal of Neuroscience*, *37*, 371–382. doi:10.1523/JNEUROSCI.2532-16.2017

Gluth, S., Spektor, M. S., & Rieskamp, J. (2017). Value-based attentional capture affects multi-alternative decision making. Manuscript under revision.

Hancock, T. O., Hess, S., & Choudhury, C. F. (2018). Decision field theory: Improvements to current methodology and comparisons with standard choice modelling techniques. *Transportation Research Part B: Methodological*, *107*, 18–40. doi:10.1016/j.trb.2017.11.004

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*, 493–518. doi:10.1002/bdm.598

Herne, K. (1999). The effects of decoy gambles on individual choice. *Experimental Economics*, *2*, 31–40. doi:10.1023/A:1009925731240

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*, 534–539. doi:10.1111/j.0956-7976.2004.00715.x

Hertwig, R. & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, *13*, 517–523. doi:10.1016/j.tics.2009.09.004

Hotaling, J. M., Busemeyer, J. R., & Li, J. (2010). Theoretical developments in decision field theory: Comment on Tsetsos, Usher, and Chater (2010). *Psychological Review*, *117*, 1294–1298. doi:10.1037/a0020401

Hu, J. & Yu, R. (2014). The neural correlates of the decoy effect in decisions. *Frontiers in Behavioral Neuroscience*, *8*, 1–8. doi:10.3389/fnbeh.2014.00271

Huber, J., Payne, J. W., & Puto, C. P. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*, 90–98. doi:10.1086/208899

Huber, J., Payne, J. W., & Puto, C. P. (2014). Let's be honest about the attraction effect. *Journal of Marketing Research*, *51*, 520–525. doi:10.1509/jmr.14.0208

Hunt, L. T., Dolan, R. J., & Behrens, T. E. J. (2014). Hierarchical competitions subserving multi-attribute choice. *Nature Neuroscience*, *17*, 1–14. doi:10.1038/nn.3836

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–292. doi:10.2307/1914185

Klein, T. A., Ullsperger, M., & Jocham, G. (2017). Learning relative values in the striatum induces violations of normative decision making. *Nature Communications*, *8*, 16033. doi:10.1038/ncomms16033

Knox, W. B., Otto, A. R., Stone, P., & Love, B. C. (2012). The nature of belief-directed exploratory choice in human decision-making. *Frontiers in Psychology*, *2*, 1–12. doi:10.3389/fpsyg.2011.00398

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*, 1292–1298. doi:10.1038/nn.2635

Kreps, D. M. (1990). *A Course in Microeconomic Theory*. Princeton, NJ: Princeton University Press.

Lichtenstein, S. & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*, 46–55. doi:10.1037/h0031207

Liew, S. X., Howe, P. D. L., & Little, D. R. (2016). The appropriacy of averaging in the study of context effects. *Psychonomic Bulletin & Review*, *23*, 1639–1646. doi:10.3758/s13423-016-1032-7

Lim, S.-L., O'Doherty, J. P., & Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *Journal of Neuroscience*, *31*, 13214–13223. doi:10.1523/JNEUROSCI.1246-11.2011

Louie, K., Grattan, L. E., & Glimcher, P. W. (2011). Reward value-based gain control: Divisive normalization in parietal cortex. *Journal of Neuroscience*, *31*, 10627–10639. doi:10.1523/JNEUROSCI.1237-11.2011

Louie, K., Khaw, M. W., & Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, *110*, 6139–6144. doi:10.1073/pnas.1217854110

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York, NY: Wiley.

Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2014). Extreme outcomes sway risky decisions from experience. *Journal of Behavioral Decision Making*, *27*, 146–156. doi:10.1002/bdm.1792

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin & Review*, *21*, 629–636. doi:10.3758/s13423-013-0542-9

Malkoc, S. A., Hedgcock, W., & Hoeffler, S. (2013). Between a rock and a hard place: The failure of the attraction effect among unattractive alternatives.

Journal of Consumer Psychology, *23*, 317–329. doi:10.1016/j.jcps.2012.10.008

Mohr, P. N. C., Heekeren, H. R., & Rieskamp, J. (2017). Attraction effect in risky choice can be explained by subjective distance between choice alternatives. *Scientific Reports*, *7*, 8942. doi:10.1038/s41598-017-06968-5

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, *30*, 12366–12378. doi:10.1523/JNEUROSCI.0822-10.2010

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience*, *32*, 551–562. doi:10.1523/JNEUROSCI.5498-10.2012

Noguchi, T. & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition*, *132*, 44–56. doi:10.1016/j.cognition.2014.03.006

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, *6*, 8096. doi:10.1038/ncomms9096

Parrish, A. E., Evans, T. A., & Beran, M. J. (2015). Rhesus macaques (*Macaca mulatta*) exhibit the decoy effect in a perceptual discrimination task. *Attention, Perception, & Psychophysics*, *77*, 1715–1725. doi:10.3758/s13414-015-0885-6

Pettibone, J. C. (2012). Testing the effect of time pressure on asymmetric dominance and compromise decoys in choice. *Judgment and Decision Making*, *7*, 513–521. Retrieved from http://journal.sjdm.org/11/111114/jdm111114.pdf

Pettibone, J. C. & Wedell, D. H. (2000). Examining models of nondominated decoy effects across judgment and choice. *Organizational Behavior and Human Decision Processes*, *81*, 300–328. doi:10.1006/obhd.1999.2880

Plonsky, O. & Erev, I. (2017). Learning in settings with partial feedback and the wavy recency effect of rare events. *Cognitive Psychology*, *93*, 18–43. doi:10.1016/j.cogpsych.2017.01.002

Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review*, *122*, 621–647. doi:10.1037/a0039413

Rakow, T., Newell, B. R., & Wright, L. (2015). Forgone but not forgotten: The effects of partial and full feedback in "harsh" and "kind" environments. *Psychonomic Bulletin & Review*, *22*, 1807–1813. doi:10.3758/s13423-015-0848-x

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108. doi:10.1037/0033-295X.85.2.59

Ratcliff, R. & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367. doi:10.1037/0033-295X.111.2.333

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, *44*, 631–661. doi:10.1257/jel.44.3.631

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review*, *108*, 370–392. doi:10.1037/0033-295X.108.2.370

Savage, L. J. (1954). *The Foundations of Statistics*. New York, NY: Wiley.

Scarpi, D. (2011). The impact of phantom decoys on choices in cats. *Animal Cognition*, *14*, 127–136. doi:10.1007/s10071-010-0350-9

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599. doi:10.1126/science.275.5306.1593

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*, 1317–1322. doi:10.1038/nn1150

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*, 158–174. doi:10.1086/209205

Simonson, I. (2014). Vices and virtues of misguided replications: The case of asymmetric dominance. *Journal of Marketing Research*, *51*, 514–519. doi:10.1509/jmr.14.0093

Simonson, I. & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, *29*, 281–295. doi:Doi10.2307/3172740

Soltani, A., De Martino, B., & Camerer, C. (2012). A range-normalization model of context-dependent choice: A new model and evidence. *PLoS Computational Biology*, *8*, 1–15. doi:10.1371/journal.pcbi.1002607

Speekenbrink, M. & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science*, *7*, 351–367. doi:10.1111/tops.12145

Spektor, M. S., Gluth, S., Fontanesi, L., & Rieskamp, J. (2017). How similarity between choice options affects decisions from experience: The accentuation of differences model. Manuscript under revision.

Spektor, M. S., Kellen, D., & Hotaling, J. M. (2017). When the good looks bad: An experimental exploration of the repulsion effect. Manuscript under revision.

Steingroever, H., Wetzels, R., & Wagenmakers, E.-J. (2013). A comparison of reinforcement learning models for the Iowa gambling task using parameter space partitioning. *Journal of Problem Solving*, *5*, 1–32. doi:10.7771/1932-6246.1150

Steingroever, H., Wetzels, R., & Wagenmakers, E.-J. (2014). Absolute performance of reinforcement-learning models for the Iowa gambling task. *Decision*, *1*, 161–183. doi:10.1037/dec0000005

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology*, *53*, 1–26. doi:10.1016/j.cogpsych.2005.10.003

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662. doi:10.1037/h0054651

Sutton, R. S. & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tan, K., Dong, S., Liu, X., Chen, W., Wang, Y., Oldroyd, B. P., & Latty, T. (2015). Phantom alternatives influence food preferences in the eastern honeybee *Apis cerana*. *Journal of Animal Ecology*, *84*, 509–517. doi:10.1111/1365-2656.12288

Trueblood, J. S. (2012). Multialternative context effects obtained using an inference task. *Psychonomic Bulletin & Review*, *19*, 962–968. doi:10.3758/s13423-012-0288-9

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, *121*, 179–205. doi:10.1037/a0036137

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2015). The fragile nature of contextual preference reversals: Reply to Tsetsos, Chater, and Usher (2015). *Psychological Review*, *122*, 848–853. doi:10.1037/a0039656

Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers: Context effects are fundamental to decision making. *Psychological Science*, *24*, 901–908. doi:10.1177/0956797612464241

Trueblood, J. S. & Dasari, A. (2017). The impact of presentation order on the attraction effect in decision-making. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th annual conference of the cognitive science society* (pp. 3374–3379). Austin, TX: Cognitive Science Society.

Trueblood, J. S. & Pettibone, J. C. (2017). The phantom decoy effect in perceptual decision making. *Journal of Behavioral Decision Making*, *30*, 157–167. doi:10.1002/bdm.1930

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, *109*, 9659–9664. doi:10.1073/pnas.1119569109

Tsetsos, K., Chater, N., & Usher, M. (2015). Examining the mechanisms underlying contextual preference reversal: Comment on Trueblood, Brown, and Heathcote (2014). *Psychological Review*, *122*, 838–847. doi:10.1037/a0038953

Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., & Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences*, *113*, 3102–3107. doi:10.1073/pnas.1519157113

Tsetsos, K., Usher, M., & Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological Review*, *117*, 1275–1293. doi:10 . 1037 / a0020580

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*, 31–48. doi:10.1037/h0026750

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299. doi:10.1037/h0032955

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458. doi:10.1126/science.7455683

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323. doi:10.1007/BF00122574

Usher, M. & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592. doi:10.1037/0033-295X.108.3.550

Usher, M. & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*, 757–769. doi:10.1037/0033-295X.111.3.757

Vlaev, I., Chater, N., Stewart, N., & Brown, G. D. (2011). Does the brain calculate value? *Trends in Cognitive Sciences*, *15*, 546–554. doi:10.1016/j.tics.2011.09.008

von Neumann, J. & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Wedell, D. H. (1991). Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 767–778. doi:10.1037//0278-7393.17.4.767

Wollschläger, L. M. & Diederich, A. (2012). The $2N$-ary choice tree model for N-alternative preferential choice. *Frontiers in Psychology*, *3*, 1–11. doi:10.3389/fpsyg.2012.00189

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2017). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*. Advance online publication. doi:10.1037/bul0000115

Yang, S. & Lynn, M. (2014). More Evidence Challenging the Robustness and Usefulness of the Attraction Effect. *Journal of Marketing Research*, *51*, 508–513. doi:10.1509/jmr.14.0020

Yechiam, E. & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review*, *12*, 387–402. doi:10.3758/BF03193783

Yechiam, E. & Busemeyer, J. R. (2006). The effect of foregone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making*, *19*, 1–16. doi:10.1002/bdm.509

Yechiam, E. & Busemeyer, J. R. (2008). Evaluating generalizability and parameter consistency in learning models. *Games and Economic Behavior*, *63*, 370–394. doi:10.1016/j.geb.2007.08.011

Yechiam, E., Stout, J. C., Busemeyer, J. R., Rock, S. L., & Finn, P. R. (2005). Individual differences in the response to forgone payoffs: An examination of high functioning drug abusers. *Journal of Behavioral Decision Making*, *18*, 97–110. doi:10.1002/bdm.487

Zeigenfuse, M. D., Pleskac, T. J., & Liu, T. (2014). Rapid decisions from experience. *Cognition*, *131*, 181–194. doi:10.1016/j.cognition.2013.12.012

Appendix A

Spektor, Gluth, Fontanesi, and Rieskamp (2017)

How similarity between choice options affects decisions from experience: The accentuation of differences model

Mikhail S. Spektor, Sebastian Gluth, Laura Fontanesi, and Jörg Rieskamp
University of Basel

Author Note

Mikhail S. Spektor, Sebastian Gluth, Laura Fontanesi, and Jörg Rieskamp, Faculty of Psychology, University of Basel, Switzerland.

This research was supported by the Swiss National Science Foundation (SNSF Grant 100014_153616). We thank Anita Todd for editing the manuscript.

Corresponding author: Mikhail S. Spektor, Faculty of Psychology, Missionsstrasse 62a, 4055, Basel, Switzerland.

Email: michael@spektor.ch

Abstract

Traditional theories of decision making typically assume that humans evaluate choice options independently of each other. The independence principle underlying this notion states that the relative choice probability of two options should be independent of the choice set. Previous research demonstrated systematic violations of this principle in decisions from description (i.e., context effects), yet it remains unclear whether these effects also occur in decisions from experience. Existing reinforcement learning models describing experience-based decisions do not predict context effects. This study provides both experimental evidence for context effects in decisions from experience and a psychologically motivated reinforcement learning model that explains the behavioral phenomena. In three experiments, the similarity effect, compromise effect, and attraction effect were explored in a 3-armed bandit task with full feedback. Participants' behavior systematically violated the independence principle, although mostly not in line with past context-effect patterns. The observed similarity effect and the reversals of the compromise effect and the attraction effect can be explained by a similarity mechanism according to which options with similar outcomes inhibit each other, making them appear less attractive. We propose the accentuation of differences model that relies on this mechanism. It outperforms traditional reinforcement learning models in describing the observed findings. Its generalizability was successfully tested using an independent data set examining learning in the Iowa gambling task. In sum, the present work is the first to demonstrate systematic violations of the independence principle in decisions from experience and offer a psychologically motivated model to explain the observed phenomena.

Keywords: context effects, similarity effect, reinforcement learning, decisions from experience, decision making

Many decisions in everyday life are made on the basis of experience: People repeatedly choose between different brands of chocolate until they find the one that suits their taste best, or they try different types and doses of painkillers to relieve their headaches. Likewise, they try out different workout programs to find the right level of exercise. These decisions from experience differ from decisions from descriptions, in which people choose between alternatives on the basis of explicit verbal descriptions of the decision alternatives. Research on the differences between the two decision types gained prominence with the discovery of the *description–experience gap in risky choice* (see Hertwig & Erev, 2009). This phenomenon hints at the possibility that there are fundamentally different processes involved in decisions from experience and those from description. The present paper explores one of the phenomena known from decisions from description, *context effects*, in a decision-from-experience setting. We demonstrate that there is a significant mismatch between the two decision types and provide a novel cognitive model to explain this new "gap."

Many economic choice theories require the *independence of irrelevant alternatives* (IIA) principle, according to which the relative choice probability of two options should be independent of the choice set (Rieskamp, Busemeyer, & Mellers, 2006). Context effects in preferential choice, on the other hand, demonstrate that this choice probability can be systematically altered, by either adding or removing specific choice options to/from the choice set. So far, these effects have only been demonstrated in decisions from description. Interestingly, models of experience-based decision making also require that behavior obeys the IIA principle. In decisions from description, which context effect arises appears to depend on a precise placement of options in the multiattribute space. In general, similar options tend to take away the choice share of options close by (Tversky, 1972). However, if one of the similar options dominates the other, then the choice share of the dominant option can increase (Huber, Payne, & Puto, 1982; Pettibone & Wedell, 2000). Critically, for these dominance-driven IIA violations, dominance needs to be clearly perceived (cf. Huber, Payne, & Puto, 2014, p. 522 f.).

We argue that this perception is attenuated in decisions from experience, where

attribute values are not explicitly stated but have to be learned over many trials. In particular, the fact that participants only see feedback about single outcomes instead of full descriptions of the outcome distributions makes such comparisons difficult. However, comparing the single outcomes of different options with each other is quite easy.

We propose a new learning model that incorporates the idea that individuals always compare outcomes of different options with each other. This learning model combines a standard *reinforcement learning* (RL; Sutton & Barto, 1998) model with an outcome-comparison process. In this comparison process, options with similar feedback inhibit each other, rendering them less attractive, while options that stand out receive more attention. This new model, the *accentuation of differences* (AOD) model, is tested in a decisions-from-experience paradigm designed to investigate whether context effects that have been found in decisions-from-description research can also be observed in experience-based decisions.

Foreshadowing our results, the context played a vital role in shaping our participants' preferences. Yet, the context effects we observed do not match the effects observed in description-based decisions. Most importantly, standard RL models requiring IIA cannot explain the observed violations, whereas the AOD model is able to describe the data accurately.

Context Effects in Preferential Choice

Economic choice theories often require that human behavior fulfills certain consistency principles (e.g., Luce, 1959; von Neumann & Morgenstern, 1947). One major principle required by many theories is IIA, which states that the choice probability of one option relative to another option is independent of the choice set in which the options are presented (see Luce, 1959, p. 9). Shortly after the emergence of Luce's choice axiom (1959) that assumes IIA, Becker, Degroot, and Marschak (1963) provided the first evidence of a phenomenon that later became known as the *similarity effect*. The similarity effect states that adding an option to a choice set "hurts" similar

alternatives more than dissimilar ones (see Tversky, 1972, p. 283), thereby representing a violation of IIA.

Tversky (1972) developed a model to explain this phenomenon and put forward his similarity hypothesis. According to this theory, similar options (i.e., options sharing many aspects) are more likely to be jointly eliminated or jointly kept in the choice set than dissimilar ones, producing the similarity effect. To challenge Tversky's similarity hypothesis, Huber et al. (1982) proposed a mechanism based on *dominance* (i.e., one option being better on at least one attribute while being at least equally good on all other attributes) and observed a phenomenon that has been called the *asymmetric dominance* or *attraction* effect: With a choice set consisting of two options, when an asymmetrically dominated third option (i.e., an option that is dominated by one of the options but not the other) is added, the dominating option becomes more attractive relative to the other original option. The authors proposed two main explanations for it: either a decrease of sensitivity on one attribute dimension (range explanation) or an increase of importance of one attribute dimension (frequency explanation). In both cases, options' attributes are compared and trade off against each other. The attraction effect violates not only IIA but also the regularity principle, according to which the absolute choice probability for one option should not increase when additional choice options are added to the choice set (cf. Rieskamp et al., 2006). In the present paper we focus only on violations of IIA and not regularity. Similar to the mechanisms discussed for the attraction effect is Simonson's (1989) verbal theory of attributewise comparisons in which humans seek justifiable reasons for decisions. Simonson predicted a new phenomenon—the *compromise effect*. According to this effect, options that form compromises between other, more extreme options are preferred to their equally attractive extreme counterparts.

Although other context effects have been discussed in the literature, such as the *phantom decoy effect* (e.g., Doyle, O'Connor, Reynolds, & Bottomley, 1999; Trueblood & Pettibone, 2017) and the *single-option aversion* (Mochon, 2013), the triad of similarity effect, attraction effect, and compromise effect has received the most

attention (for an overview, see Rieskamp et al., 2006). Context effects have been demonstrated across many different domains and modalities. They have been reported mainly in consumer choices (e.g., Berkowitsch, Scheibehenne, & Rieskamp, 2014; Doyle et al., 1999; Heath & Chatterjee, 1995; Huber et al., 1982; Noguchi & Stewart, 2014; Pettibone, 2012) but also in lottery decisions (Herne, 1999; Hu & Yu, 2014; Soltani, De Martino, & Camerer, 2012; Tversky, 1972; Wedell, 1991) and intertemporal choices (Gluth, Hotaling, & Rieskamp, 2017) and by using inferential (Trueblood, 2012) and perceptual tasks (Trueblood, Brown, Heathcote, & Busemeyer, 2013; Zeigenfuse, Pleskac, & Liu, 2014). With the broad scope of domains as well as demonstrations with other mammalian and nonmammalian animals (Parrish, Evans, & Beran, 2015; Scarpi, 2011; Tan et al., 2015), context effects appear to be important and universal phenomena of decision making.

Reinforcement Learning and Decisions From Experience

The arguably most common paradigms used in decision-making research are based on decisions from description, where people make decisions on the basis of full descriptions of the choice alternatives and the environment. In contrast to this research tradition, paradigms using decisions from experience only provide information about the identity of the choice alternatives, while all other properties about the alternatives, as well as the environment, have to be inferred from the observations made. A very popular paradigm for studying decisions from experience is the *Iowa gambling task* (IGT; Bechara, Damasio, Damasio, & Anderson, 1994). The IGT has been used to study both decision-making deficiencies in clinical populations (Brevers, Bechara, Cleeremans, & Noël, 2013) and decision-making processes in cognitive psychology (Steingroever, Wetzels, & Wagenmakers, 2013; Wetzels, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2010). In the IGT, participants repeatedly choose one of four decks of cards. The selected deck produces one outcome from the deck's outcome distribution and a new trial begins. With such a structure, the IGT essentially corresponds to a variation of the n-armed bandit problem (Sutton & Barto, 1998) with $n = 4$.

This problem is a typical case of RL, a subdiscipline of machine learning with the aim of developing models that optimize behavior by a trial-and-error learning process. Essentially, RL models form expectations about the choice alternatives upon which the decisions are based and update them with incoming experience. One of the most common classes of RL models, temporal-difference learning, updates expectations by using the *reward prediction error* (RPE), which is the difference between the obtained and the expected reward. With the rise in popularity of the IGT and discovery of a neural substrate of the RPE in the primate brain (Schultz, Dayan, & Montague, 1997), the use of RL models has increased dramatically.

More recently, the discovery of the *description–experience gap in risky choice* (Hertwig, Barron, Weber, & Erev, 2004; Hertwig & Erev, 2009) led to increased interest in differences between decisions from description and decisions from experience. This gap is the observation that the typical decisions-from-description finding of overweighting of rare events (e.g., Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) reverses in decisions from experience (i.e., people act as if they underweight rare events). This finding raises the important question of whether the evaluation of choice alternatives might be fundamentally different in the two domains. Therefore, it is an open question whether context effects observed for decisions from description generalize to decisions from experience.

Modeling Challenges and Open Questions

Context effects are a serious challenge for standard utility models. To account for the phenomena, a variety of novel process models of decision making has been proposed (e.g., Bhatia, 2013; Roe, Busemeyer, & Townsend, 2001; Trueblood, Brown, & Heathcote, 2014; Usher & McClelland, 2004; Wollschläger & Diederich, 2012). Some contemporary RL models have adapted mechanisms of descriptive decision-making models, such as risk aversion (e.g., Niv, Edlund, Dayan, & O'Doherty, 2012; Yechiam & Busemeyer, 2005, 2008) or loss aversion (e.g., Steingroever et al., 2013, 2014; Yechiam & Busemeyer, 2005). Yet, so far no RL model has been developed that incorporates

mechanisms that could explain context effects in decisions from experience. Likewise, models that have been proposed to explain specific effects for experienced-based decisions (Erev & Haruvy, 2005; Erev & Roth, 2014; or Plonsky & Erev, 2017; Plonsky, Teodorescu, & Erev, 2015) do not address potential context effects.

In decisions from description, current models assume that attributes are explicitly represented and that they trade off and interact with each other. In such a situation, small changes of attribute values are easily perceived, potentially leading to large preference changes (see Bhatia, 2013; Trueblood et al., 2014; Tsetsos, Usher, & Chater, 2010). In decisions from experience, on the other hand, there are no evident attributes to observe, as options are merely characterized by the outcomes they yield. While it certainly is possible to form an internal representation of options that is based on multiple attributes, such as expected value and variability of outcomes (e.g., Mikhael & Bogacz, 2016), it is cognitively demanding and unclear in what way the attributes might interact with each other.

The aim of the present study was to explore violations of IIA in decisions from experience and thereby test a psychologically motivated RL model. We introduced a similarity mechanism into a common RL model. We present a series of three experiments in which the options were aligned in such a way that the similarity effect, the compromise effect, and the attraction effect, respectively, were expected to result. We rigorously compared our new model to other RL model variants in our experiments and tested its generalizability in an independent data set that examined learning in the IGT.

Cognitive Models

To understand the cognitive processes underlying behavior, we used four learning models and compared them against each other: three RL models that have been proposed in previous work (that all obey IIA) and one newly developed RL model that predicts violations of IIA. We denote this model the AOD model, referring to the new similarity mechanism that accentuates differences between negatively correlated

options. The three other models include a standard fixed-learning-rate RL model (FLR) and two risk-attitude additions to FLR: a risk-preference RL model (RPR) and a risk-sensitive RL model (RSE).

Standard RL Models

FLR model. The subjective expectations $X_{i,t}$ of option i for the trial following the decision $t+1$ are updated according to a RPE-based updating mechanism with a fixed learning rate:

$$X_{i,t+1} = X_{i,t} + \alpha\Big(s(O_{i,t}) - X_{i,t}\Big), \tag{1}$$

where $s(O_{i,t})$ is the subjective value of the outcome observed on trial t. In this case, $s(O_{i,t}) = O_{i,t}$, where $O_{i,t}$ is the outcome observed on trial t. FLR has one free parameter, learning rate α, ranging from 0 to 1. With a learning rate of $\alpha = 0$, there is no updating and $X_{i,t+1} = X_{i,t}$. With a learning rate of $\alpha = 1$, there is a perfect recency effect and $X_{i,t+1} = s(O_{i,t})$. The FLR model is the most basic within the RL framework, and it predicts neither risk preferences nor context effects. The only free parameter in this model is the degree to which participants adapt to recently observed outcomes.

RPR model. The RPR model is a more general version of the FLR that captures risk preferences by assuming a marginal utility function that maps observed rewards to subjective representations using a power-utility function. As such, $s(O_{i,t})$ is now a function of the observed outcome $f(O_{i,t})$, where

$$s(O_{i,t}) = f(O_{i,t}) = \begin{cases} O_{i,t}^{\gamma} & \text{if} \quad O_{i,t} \geq 0 \\ -|O_{i,t}|^{\gamma} & \text{if} \quad O_{i,t} < 0. \end{cases} \tag{2}$$

The FLR model is a special case of the RPR model in which $\gamma = 1$, representing risk-neutral behavior. With $\gamma > 1$, risk-seeking behavior is represented, and $0 < \gamma < 1$ represents risk-averse behavior. The RPR model is thus a generalization of the utility model from Niv et al. (2012) for more than two outcomes. Risk attitude is an important concept necessary for the understanding of empirical phenomena as many people

exhibit risk aversion. As such, models incorporating risk preferences often outperform those that do not (e.g., Gershman, 2015; Niv et al., 2012).

RSE model. Similar to the RPR model, the RSE model captures risk preference during the expectation-updating phase. It has been successfully used in the literature before (e.g., Frank, Doll, Oas-Terpstra, & Moreno, 2009; Niv et al., 2012). Instead of forming subjective reward as a function of the objective reward, the RSE model assumes different learning rates depending on the RPE:

$$\alpha = \begin{cases} \alpha^{+} & \text{if} \quad s(O_{i,t}) - X_{i,t} \geq 0 \\ \alpha^{-} & \text{if} \quad s(O_{i,t}) - X_{i,t} < 0. \end{cases} \tag{3}$$

If $\alpha^{+} > \alpha^{-}$, positive RPEs have a higher impact on expectation updating and thus risk-seeking behavior is represented, and if $\alpha^{+} < \alpha^{-}$, risk-averse behavior is represented. RSE is identical to FLR if $\alpha^{+} = \alpha^{-}$ and $s(O_{i,t}) = O_{i,t}$, the latter of which we assume.

The AOD Model

A core component of explaining context effects in many prominent process models for description-based decisions is an attention-switching-between-attributes mechanism (e.g., Tversky, 1972). Within a single trial, attention repeatedly shifts between the options' attributes, and the options are compared to each other on the currently attended attribute. During this comparison process, the options interact with each other in different ways. For instance, they inhibit each other (Roe et al., 2001), or disadvantageous comparisons are more prominently weighted (Usher & McClelland, 2004), similar to loss aversion (Kahneman & Tversky, 1979). As discussed earlier, in an experienced-based decision problem under uncertainty, individuals have to *learn* the outcome distributions of the different options. We do not assume that they acquire an explicit representation of different attributes. Instead, we assume a dynamic learning process during which people compare the feedback they receive about the options with each other.

Our approach, albeit operating within the general framework of RL, crucially

differs from existing models in the literature in an important aspect. While existing models assume that the subjective expectations of all options are updated independently of each other, we expect options to interact with each other during the learning process. We propose that the similarity of the different options' outcomes has an inhibitory effect on the evaluation of the options. More specifically, when two options have similar outcomes, these options inhibit each other so that they are perceived as less attractive. In contrast to the lateral-inhibition mechanism introduced by Roe et al. (2001), our *similarity mechanism* is sign independent. Comparisons involving close and inferior options do not increase the attractiveness of the respective superior options. This similarity mechanism can also be interpreted as an attentional mechanism, where options with outcomes that are dissimilar to other options stand out and therefore receive more attention and are consequently perceived as more attractive (e.g., Krajbich, Armel, & Rangel, 2010).

It is important to stress that our approach assumes that the similarity mechanism is active during the learning process and results in a single subjective value for each option on each trial. During choice, these values do not interact with each other. Thereby our approach is similar to Trueblood et al.'s (2014) decomposition of choice into a front-end process (formation of preferences) and a back-end process (choice/error model) and contrasts with most existing models in which these two processes are intertwined (e.g., Roe et al., 2001; Usher & McClelland, 2004).

We formalize the similarity mechanism in the AOD as follows. The subjective value of the outcome of option i is defined as

$$s(O_{i,t}) = f(O_{i,t}) - \eta \times S_{i,t} \times \overline{|f(O_{J,z})|}. \tag{4}$$

where $f(O_{i,t})$ comes from Equation 2, $S_{i,t}$ is option i's average similarity with all other options, η is the parameter that governs the balance between subjective utility and the similarity-based inhibition, and $\overline{|f(O_{J,z})|}$ is the sum of unsigned subjective utilities of all options J divided by the number of options. Subscript z corresponds to the most recent trial on which the outcome of the individual options has been observed. If for any option

no outcomes have been observed, then the initial value $X_0 = 0$ is used. In the case of a learning situation in which the outcomes of all available options can be observed (i.e., full feedback), the most recent trial is the current trial for all options, $z = t$. Following the literature on exemplar models (see Nosofsky & Johansen, 2000), we define $S_{i,t}$ as the average similarity between option i and all other options using the city-block distance,

$$S_{i,t} = \frac{\sum_{j=1}^{J \setminus i} e^{-\psi \times |f(O_{i,z}) - f(O_{j,z})|}}{J - 1}. \tag{5}$$

Here, parameter ψ serves as a scaling parameter for the sensitivity to subjective reward differences. The average similarity $S_{i,t}$ ranges from 0 (all options are very distinct in their outcomes and/or sensitivity parameter ψ is very high) to 1 (all options are the same and/or sensitivity parameter ψ is very low). Note that parameter ψ and $\overline{|f(O_{J,z})|}$ in Equation 4 serve scaling purposes and are, within the scope of the present work, of little interest.

The main parameter of interest, η, controls the degree to which close options inhibit (or boost) each other. For $\eta > 0$, the standard case we assume, close options inhibit each other more strongly than distant options. In such a case, the model generates a preference for options that have high expected values (left term of Equation 4) and are far away from other options (right term of Equation 4), giving rise to the similarity effect but inhibiting the attraction effect. In case of the predicted similarity effect and the inhibited attraction effect, similar alternatives are perceived as less attractive compared to dissimilar options. The compromise effect is inhibited as well, although for less intuitive reasons. Consider the case of lotteries that maximized the compromise effect in Herne's (1999) Experiment 2 (gamble set c): x: (100, .3, 0), y: (60, .5, 0), z_x: (150, .2, 0), and z_y: (50, .6, 0). The choice set consisting of options x, y, and z_x makes x a compromise, whereas the choice set consisting of x, y, and z_y makes y a compromise. Simulating outcomes from these options' outcome distributions (assuming $\psi = 0.1$) results in total average similarity of .46 and .38 for x and y, respectively, in the former choice set, and total average similarity of .32 and .33 for x

and y, respectively, in the latter. Beyond the expected utility of the options (left term of Equation 4), the preferences are strongly influenced by this total average similarity. The similarities lead to a relative preference of x over y when z_y is present, and a relative preference of y over x when z_x is present. These predictions run counter to the compromise effect observed by Herne (1999) and predict the *reversal* of a compromise effect. However, these predictions only hold when $\eta > 0$; they reverse when $\eta < 0$: Closer options boost each other, giving rise to the attraction effect and compromise effect, but inhibiting the similarity effect. Note that the prediction of all three context effects not necessarily arising within the same participant are shared with other cognitive models (e.g., Wollschläger & Diederich, 2012; cf. Tsetsos, Chater, & Usher, 2015) and match previous behavioral findings in decisions from description, where the attraction effect and the compromise effect correlated positively with each other and negatively with the similarity effect (Berkowitsch et al., 2014; Trueblood, Brown, & Heathcote, 2015). If $\eta = 0$, then the AOD model reduces to the RPR model and does not predict any context-dependent preferences.

Within the AOD framework, two properties determine the closeness of options: absolute values of outcomes and their correlations. Take, for instance, options a: (10, .5, 0) and b: (20, .5, 0) and assume that $\psi = 0.1$. If the outcomes of a and b are uncorrelated, $r_{O(\text{a,b})} = 0$, then in .25 of the cases, the outcomes will be (10, 20), resulting in a distance of 10 between the options, corresponding to an average similarity of .37. All other cases (10, 0), (0, 20), and (0, 0) are equiprobable and result in average similarities of .37, .14, and 1, respectively. The total average similarity will be .47. Now imagine the outcomes of a and b are perfectly correlated, $r_{O(\text{a,b})} = 1$, so that only two outcome combinations are possible: (10, 20) and (0, 0), corresponding to the average similarities .37 and 1, respectively. The total average similarity increases to .69, compared to .47 in the uncorrelated case. Similarly, in the case of a perfectly negative correlation, $r_{O(\text{a,b})} = -1$, only the two combinations (10, 0) and (0, 20) are possible, corresponding to the average similarities .37 and .14, respectively. The total average similarity decreases to .26, compared to .47 in the uncorrelated case. To demonstrate

the influence of absolute values, take option c: (50, .5, 0) instead of b. Irrespective of its higher value, the average similarity to a is affected by the correlation of outcomes. The total average similarity would be .35, .51, and .18 for the uncorrelated, positively correlated, and negatively correlated case, respectively. Thus, correlations between risky options affect similarities, with larger similarities making similar options less or more attractive to the decision maker (for positive or negative values of η, respectively).

Note that it is possible to specify many features of the AOD model differently. For instance, perceived distance could be Euclidian or follow a Gaussian function (e.g., Hotaling, Busemeyer, & Li, 2010) instead of the proposed city-block distance (this would result in stronger predictions for the compromise effect). Also, inhibition could take another form: Instead of the observed outcomes, the subjective expectations or a mixture of both could form the basis for inhibition. In the present study we tested only the most basic variant of the model described above.

Error Model

As is common in the literature (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Niv et al., 2012; Yechiam & Busemeyer, 2005), we used a soft-max choice rule to specify the choice probabilities. It is a one-parameter error model that transforms the subjective expectation of each option $X_{i,t}$ to a choice probability $Pr(i,t)$ according to

$$Pr(i,t) = \frac{e^{\theta X_{i,t}}}{\sum_{j=1}^{J} e^{\theta X_{j,t}}}. \tag{6}$$

Choice sensitivity $\theta \geq 0$ quantifies the number of errors made. With $\theta = 0$, choices are completely random, and as $\theta \to \infty$, choices become deterministic in that the option with the highest subjective expectation is always chosen.

Goals

To explore violations of IIA in decisions from experience and test our new AOD model, we conducted three experiments. In the first experiment we aimed to test

whether the similarity effect as found in description-based choices arises in decisions from experience. The second experiment aimed to test the compromise effect, and the third experiment to test the attraction effect. According to the similarity mechanism of the AOD model, behavior in all experiments should be phenomenologically identical, whereby participants prefer rather dissimilar options. In all the experiments, we not only looked at the behavioral effects but also rigorously compared the four cognitive models against each other. As a further test of the AOD model, we used the IGT, a task not specifically designed to test context effects, as the basis for model comparison.

Experiment 1

In the first experiment, we explored the similarity effect in a decisions-from-experience setting. We used a design in which participants chose between three options, where two of the options were similar to each other while at the same time dissimilar to the remaining one. For details on option placement see Design. In accordance with the literature on context effects in decisions from description, we expected the similarity effect to arise; close options should inhibit each other, as predicted by the similarity mechanism, thus increasing the preference for dissimilar options.

Method

Participants and procedure. A total of 24 participants (15 female, age: 21–45 years, $M = 25.71$ years, $SD = 5.53$) completed Experiment 1. The experiment consisted of three parts. In the first part, after giving informed consent and reading the instructions, participants went through a training phase, which allowed them to familiarize themselves with the task and allowed us to assess learning performance. In the second part, participants completed two blocks of the experimental task in a counter-balanced order. After completing the experimental task, participants were asked to report explicit judgments about the choice options they encountered in the experimental task. Afterward, the accumulated points from the training and experimental phases were converted to Swiss francs (CHF) and participants received

their bonus payment in addition to the show-up fee (CHF 20 or a course-credit equivalent of one hour per hour or part thereof) in cash. The total duration of the experiment was 25–35 min. All our experiments were programmed in `expyriment` (Krause & Lindemann, 2014) and approved by the institutional review board of the Department of Psychology, University of Basel.

Paradigm. The paradigm used in the training and experimental phases was a variation of the n-armed bandit problem (Sutton & Barto, 1998) in which participants repeatedly chose between n different options. The general payoff structure of all options was similar to a binary lottery. The outcome $O_{i,t}$ of option i on trial t yielded some outcome with a probability of π_i, or 0 otherwise. The outcome was drawn from a discretized normal distribution with mean μ_i and option-independent standard deviation σ, resulting in the gamble structure $O_{i,t} \sim \big(\mathcal{N}(\mu_i, \sigma^2), \pi_i; 0\big)$.[1] The option-specific parameters μ_i and π_i were static (i.e., did not change with time). The sequences of outcomes were unique per participant and option and were pseudorandomly generated to achieve representative observations throughout each block.

Each of the options was represented by one of eleven 5×5 identicons (horizontally symmetrical random matrices of squares) that were selected for maximum discriminability from a large pool of randomly generated ones. They were colored such that the subjective hue and saturation were very similar according to the HSL_{UV} color scheme (www.hsluv.org; see Figure A1 for complete set of stimuli used). The options were presented on-screen throughout each block. Each trial began with a fixation cross being displayed for 500–1,500 ms. After the fixation cross disappeared, participants chose one of the presented options, $i = x$. They received feedback about the outcomes of all three options $O_{\bullet,t}$ for 2,500 ms while also receiving the outcome of the chosen option $O_{x,t}$ (see Figure 1 for an example of a choice trial with three options). The accumulated points were converted to a bonus payment yielding up to CHF 6. The conversion rate was communicated to the participants prior to the experiment.

[1] Observed standard deviances differed slightly between the options due to discretization.

Design. The training phase consisted of a single block with two options and 40 trials. The options differed in their probabilities of yielding a reward π_i as well as the means of the normal distributions μ_i. A higher valued option (HV; $\mu_{\text{HV}} = 40$, $\pi_{\text{HV}} = .4$) stochastically dominated a lower valued option (LV; $\mu_{LV} = 20$, $\pi_{LV} = .2$). Both options had the same variance of the normal distributions from which the outcomes were drawn, $\sigma \approx 6.95$. The global accuracy across all 40 trials, $Acc = Pr(\text{HV})$, was the dependent measure used to assess learning performance. Lack of learning is reflected in random behavior ($Acc = .50$).

The experimental phase consisted of two blocks/choice sets with three options and 120 trials each. Choice set 1 (S_1) contained options A: ($\mu_{\text{A}} = 83$, $\pi_{\text{A}} = .17$), B: ($\mu_{\text{B}} = 78$, $\pi_{\text{B}} = .22$), and C: ($\mu_{\text{C}} = 22$, $\pi_{\text{C}} = .78$), whereas choice set 2 (S_2) contained options B, C, and D: ($\mu_{\text{D}} = 17$, $\pi_{\text{D}} = .83$). See Figure 2 for a graphical depiction of option placement. The variance was similar to that of the options in the training phase, $\sigma \approx 6.64$. Note that between the two choice sets, options B and C were represented by different stimuli (so that participants had to relearn the options' values). Options {A, B} and {C, D}, respectively, were similar to each other and dissimilar to the other two options. As such, contrasting the choice proportions between the two choice sets allowed us to measure the emergence of context effects. To do so, we used the *relative choice share of the target* (RST; Berkowitsch et al., 2014). Let $Pr(\text{Tar}|S_i)$ be the probability of choosing the target option Tar in choice set S_i and $Pr(\text{Com}|S_i)$ the probability of choosing the competitor option Com, then

$$\text{RST} = 0.5 \times \left(\frac{Pr(\text{Tar}|S_1)}{Pr(\text{Tar}|S_1) + Pr(\text{Com}|S_1)} + \frac{Pr(\text{Tar}|S_2)}{Pr(\text{Tar}|S_2) + Pr(\text{Com}|S_2)} \right). \quad (7)$$

In Experiment 1, the target option was the one option that was dissimilar to the other two options in each of the choice sets, C in S_1 and B in S_2.

RST is a measure of strong IIA, as it restricts the choice proportions to be exactly equal. In contrast to strong IIA, weak IIA does not impose this restriction. Weak IIA holds if the preference ordering remains the same. It holds if the following property is

satisfied:

$$\begin{aligned} \text{if} \quad & Pr(\text{C}|S_1) > Pr(\text{B}|S_1) \\ \text{then} \quad & Pr(\text{C}|S_2) > Pr(\text{B}|S_2). \end{aligned} \tag{8}$$

Consequently, if individuals prefer B over C in S_1, weak IIA is satisfied only if they prefer B over C in S_2 as well. Consistent preference for the target (C in S_1 and B in S_2) or the competitor (B in S_1 and C in S_2) would thus violate weak IIA. It is important to look at both violations: Indifference and noise will also lead to violations of Equation 8 without supporting context effects. Such violations would be nonsystematic, though, and would go symmetrically in both directions. We report the percentage of participants with preference reversals in both directions. Systematic violations are present if the proportion of weak IIA violations differs between the two directions.

In the third part, participants were asked to report explicit judgments about the options. The stimulus representing an option was presented at the top of the screen with four numerical scales below it. The first scale ranged from 0 to 100 and indicated the probability of winning something. The second scale ranged from ±0 to ±50 and indicated the uncertainty about the estimate of the first scale (see Figure 3 for an example). After placing an estimate on the second scale, the uncertainty was graphically represented on the first scale in the form of upper and lower bounds around the point estimate of the first scale. The third and fourth scales were similar to the first two, the only difference being the estimation of the average number of points won if an option yielded any outcome. The options were presented blockwise, starting with the chronologically last block, and randomized within blocks. To assess whether explicit knowledge of the underlying structure predicts choices, we correlated the ratings obtained from the probability judgments, the magnitude judgments, and the product of both with individual preferences.

We followed, whenever possible, a Bayesian approach for our statistical analyses (for an introduction to Bayesian data analysis see Gelman et al., 2013). To obtain the posterior distributions we used a no-U-turn sampler (Hoffman & Gelman, 2014) as

implemented in `Stan` (Carpenter et al., 2017; see Appendix for details about settings used for the sampler). Because there was no Bayesian rank-correlation variant, we conducted the analyses related to explicit judgments with frequentist statistics. All other frequentist tests led to qualitatively identical conclusions to the Bayesian variants and are provided in the Appendix.

For computational reasons, we used Gaussian prior distributions that we transformed to achieve the desired parameter space. Both behavioral measures—the training-phase accuracy *Acc* and context-effects measure RST—were restricted to the $[0, 1]$ interval, allowing us to fit a normal distribution with unknown mean μ and standard deviation σ to the probit-transformed values. As prior distributions we used a standard normal distribution $\mathcal{N}(0, 1)$ for μ (the transformation of this normal distribution back from the probit scale results in a uniform distribution between 0 and 1) and the half-normal distribution $H\mathcal{N}(0, 1.5)$ for σ (for a graphical depiction, see Figure A2). For hypothesis testing, we used the Bayesian central posterior interval (BCI; Gelman et al., 2013, p. 33 f.). The $X\%$ BCI of a parameter depicts the range in which the parameter lies with $X\%$ probability.

For the analysis of the explicit judgments, we used Kendall's τ for each option and each of three different measures (i.e., estimated magnitude, estimated probability, and their product) separately to reduce the influence of extreme estimates. As there were six unique options for each experiment and for each of the three measures, there were a total of 18 correlation coefficients in each experiment. Assuming an α-error probability of .05 for each of these tests, the likelihood of two or fewer significant results is .94 and of three or fewer significant results is .99. Across all three experiments, at least six significant correlation coefficients were necessary to reject the null hypothesis of no correlation on a 5% level.

Model comparison procedure. We implemented our model comparisons within a hierarchical Bayesian framework. We compared the models' fits using the widely applicable information criterion (WAIC; Watanabe, 2010; computed as in Vehtari, Gelman, & Gabry, 2017, using samples from the posterior distribution). We

specified the group-level distributions for each of the parameters in terms of Gaussian distributions. We used weakly informative prior distributions over the spaces of the parameters for the group-level parameters. In the case of individual-level parameters restricted within the $[0, 1]$ range (learning rates), we used the same priors as for the behavioral-measures analyses: $\mathcal{N}(0, 1)$ for μ and $H\mathcal{N}(0, 1.5)$ for σ. The individual-level parameters drawn from this distribution were then Φ transformed, resulting in the uniform prior $\mathcal{U}(0, 1)$ for the group-level mean after transformation. For individual-level parameters in the half-open interval $[0, \infty)$ (choice sensitivity θ, utility-function exponent γ, and scaling parameter ψ), we used the same prior distributions for the group-level parameters but an exponential instead of a Φ transformation. For the only parameter in our models with the parameter space $(-\infty, \infty)$, η of the AOD model, we again used the same priors but did not apply any transformations to them. See Figures A3–A6 for graphical depictions of the model specifications of FLR, RPR, AOD, and RSE, respectively.

Results

For the training phase, accuracy was assessed as the relative proportion of HV choices. The 95% BCI of this variable, *Acc*, was entirely above .5, BCI = $[.58, .77]$; $M_\mu = .68$; $SD_\mu = .05$, showing that participants were more likely to choose the HV option than the LV option, thus reflecting an ability to learn.

In the experimental phase, the similarity-effect analysis revealed that the 95% BCI of the RST was entirely above .5, BCI = $[.53, .63]$; $M_\mu = .58$; $SD_\mu = .02$. In both choice sets, participants preferred option C over the other options. The strong preference for C in S_1 became substantially weaker in S_2. The presence of D "hurt" the similar option C more than the dissimilar option B (similarity effect; see Figure 4, top left panel).

Analysis of weak IIA violations revealed that 29% of participants showed a preference reversal in line with the similarity effect: They preferred C over B in the choice set S_1 and B over C in the choice set S_2. Only 4% showed a preference reversal in the opposite direction (i.e., preference for B over C in S_1 and C over B in S_2).

To assess explicit knowledge about the options' properties, responses to the post questionnaire were analyzed as described above. Participants' ranks in the estimations of probabilities and magnitudes did not correlate with how often they chose the options (all ps $\geq .09$).

Model Comparison

The baseline model had a deviance of 12,656, meaning that models with adjusted-for-complexity deviances below this value explain the data beyond pure guessing. All of our candidate models surpassed this criterion. The AOD outperformed all the other models, having the by-far lowest WAIC (10,712) compared to the runner-up RPR (WAIC: 10,973; see Table 1 for all information criteria). The ΔWAIC (i.e., the difference of the penalized log-pointwise predictive densities) of the two best models was 261 points on the deviance scale and the standard error of the differences was 31.98, resulting in a standardized effect size of $\Delta = 8.16\sigma$, reflecting a substantially better fit of the AOD model.

A consistent result across the group-level posterior distributions of all models is the rather slight adaptation to recent outcomes, as reflected in the overall low learning rates. Instead, participants included even early observations in the formation of their expectations. Additionally, participants demonstrated risk aversion: For the two models reflecting risk attitudes in the curvature of the utility function, RPR and AOD, the 95% BCI of the group-level mean of γ was entirely below 1. In the model with separate learning rates for positive and negative RPEs, RSE, participants weighted negative RPEs higher than positives ones, again reflecting risk aversion. More importantly, inhibition played an important role in forming participants' impressions of the options: Posterior estimates of the group-level mean of η lay mostly above 0, and the 95% BCI did not include 0. See Table 2 for posterior distributions of all the models' group-level mean parameters.

Posterior Predictive Checks

We performed posterior predictive checks not only to evaluate the model fits relative to each other but also to obtain an impression of the models' absolute fits to the data. To do so, for each sample from the posterior, we generated responses from virtual participants. In a first step, we checked whether the models could reproduce the empirically observed choice proportions (Figure 4, left column). The general pattern observed in the model comparison using the WAIC is also reflected in the posterior predictive checks. The AOD model's predictions most closely matched the choice proportions observed in the experiment in both choice sets, S_1 and S_2. Interestingly, despite the built-in assumption of IIA in the RPR and the RSE, both models' inclusion of risk preferences seems to have interacted with the dynamically probabilistic nature of the task, thus mimicking to some degree IIA violations.

In a second step, we aggregated choice proportions into bins of 10 trials to see if the models correctly reflected the dynamic development of choice proportions (Figure 5, left panels). As with the other comparison techniques, the AOD model also more closely resembled the development of choice proportions. In this comparison, the superiority of the AOD was most evident. The 95% Bayesian posterior interval of the mean (PI) is narrower for most of the options, indicating less spread in the predictions derived from the posterior distribution of the parameters. Additionally, the RPR model predicted a consistent preference of D over B in choice set S_2, a pattern not observed in the data. This dynamic development, a preference of D over B in the beginning that reverses with later trials, is correctly captured by the AOD model.

Discussion

In Experiment 1, we explored the similarity effect in a decisions-from-experience setting. We used a bandit task with full feedback in which participants repeatedly chose among three options in two choice sets. Two of the options in each of the choice sets were similar to each other and dissimilar to the remaining option. We observed the similarity effect systematically violating IIA. In a rigorous comparison of the models

against each other, the AOD model clearly outperformed all other RL models in predicting the observed behavior and correctly captured the dynamic development observed in the task.

Experiment 2

Experiment 2 explored the compromise effect in a decisions-from-experience setting. We used the same procedure and paradigm as in Experiment 1 but used outcome distributions similar to those eliciting a compromise effect in decisions from description. Crucially, we did not expect a compromise effect to arise. According to the similarity mechanism, the reversal of a compromise effect should arise: The compromise option is more strongly inhibited than the other options, leading to a lower preference for the compromise option. The extreme options are less strongly inhibited, leading to a weaker impact of the similarity mechanism on the extreme options and thus making them more attractive.

Method

Participants and procedure. Initially, 24 participants took part in Experiment 2. Because of one incomplete data set, we could analyze the data of only 23 participants (19 female, age: 20–54 years, $M = 25.17$ years, $SD = 7.83$). The outcome distributions were selected to be similar to those of the decisions-from-description literature on the compromise effect. A decoy option was selected that occupied an extreme position relative to the other two options, thus making one of the two core options a compromise. Otherwise, the experimental procedure was identical to Experiment 1.

Paradigm and design. The paradigm, the training phase, and the post questionnaire were identical to those in Experiment 1. Only the choice sets in the experimental phase were different. Choice set 1 (S_1) contained options A: ($\mu_A = 83$, $\pi_A = .17$), B: ($\mu_B = 61$, $\pi_B = .39$), and C: ($\mu_C = 39$, $\pi_C = .61$), whereas choice set 2 (S_2) contained options B, C, and D: ($\mu_D = 17$, $\pi_D = .83$). The variance was similar to the options in the training phase and Experiment 1, $\sigma \approx 6.56$. Between the two choice sets,

options B and C were represented by different stimuli (so that participants had to relearn the options' values). Compared to Experiment 1, options B and C were obtained by linearly filling the probability-magnitude space between A and D. This resulted in a higher expected value of options B and C than of A and D.[2] Contrasting the choice proportions between the two choice sets measures a potential context effect. To do so, we again used the RST. For reasons of coherence with the literature on decisions from description, we defined the compromise option as the target in each of the choice sets, B in S_1 and C in S_2, despite our model predicting the reversal of it.

Results

For the training phase, accuracy was assessed as the relative proportion of HV choices. The 95% BCI of this variable, *Acc*, was entirely above .5, BCI = $[.71, .88]$; $M_\mu = .81$; $SD_\mu = .04$, showing that participants were more likely to choose the HV option than the LV one, thus reflecting an ability to learn.

In the experimental phase, the compromise-effect analysis revealed that the 95% BCI of the RST was entirely below .5, BCI = $[.36, .49]$; $M_\mu = .42$; $SD_\mu = .03$. This indicates a *reversed* compromise effect: Participants had an absolute preference for C over B in choice set S_1, and an absolute preference for B over C in choice set S_2 (Figure 4, top middle panel).

This strong absolute preference reversal is reflected in the analysis of weak IIA violations. None of the participants showed a preference reversal consistent with a compromise effect (preference for B over C in S_1, and C over B in S_2). Instead, 43% showed a preference reversal in the opposite direction, reflecting the group-level average. This pattern represents a violation of IIA in the direction opposite the compromise effect.

The analysis of the post questionnaire about the options' properties did not

[2] We chose this to prevent the following situation. Imagine options a = (\$60, .4, 0) and b = (\$40, .6, 0) with the same expected value. The decoy option c for a compromise effect would have to be c_a = (\$120, .2, 0) and c_b = (\$30, .8, 0) for options a and b, respectively, to be shifted by .2 on the probability scale while still maintaining the same expected value. Such an expected-value-preserving decoy placement yields qualitatively different decoys. On the other hand, c_a = (\$80, .2, 0) and c_b = (\$20, .8, 0) results in comparable decoys, despite the expected value difference (24 vs. 16).

support the notion of explicit knowledge. Except for one case in which the estimated probability of option B in choice set S_1 correlated with the corresponding choice proportion ($r_\theta = 0.30$, $p = .047$), participants' ranks in the estimations of probabilities and magnitudes did not correlate with how often they chose the options (all other ps $\geq .15$).

Model Comparison

In Experiment 2, the baseline model had a deviance of 12,129, meaning that models with adjusted-for-complexity deviances below that threshold explained the data beyond pure guessing. All of our candidate models surpassed this criterion. The AOD model again outperformed all the other models, having the by-far lowest WAIC (10,072) compared to the runner-up RPR (WAIC: 10,514; see Table 1 for all information criteria). The ΔWAIC of the two best models was 442 points on the deviance scale and the standard error of the differences was 35.09, resulting in a standardized effect size of $\Delta = 12.59\sigma$, reflecting a substantially better fit of the AOD model.

A consistent result across the group-level posterior distributions of all models is the rather slight adaptation to recent outcomes, as reflected in the overall low learning rates. Instead, participants included even early observations in the formation of their expectations. Additionally, participants demonstrated slight risk seeking. In the two models reflecting risk attitudes in the curvature of the utility function, RPR and AOD, estimates of γ were mostly above 1. In the model with separate learning rates for positive and negative RPEs, RSE, participants weighted positive RPEs higher than negative ones, again reflecting risk seeking. Inhibition played an important role in forming participants' impressions of the options: Posterior estimates of the group-level mean of η lay mostly above 0, even though the 95% BCI included 0. See Table 3 for posterior distributions of all the models' group-level mean parameters.

Posterior Predictive Checks

We performed posterior predictive checks not only to evaluate the model fits relative to each other but also to obtain an impression of the models' absolute fits to

the data. We used the same procedure as in Experiment 1. In the first step, we checked whether the models could reproduce the empirically observed choice proportions (Figure 4, middle column). The general pattern observed in the model comparison using the WAIC was also reflected in the posterior predictive checks. The AOD model's predictions most closely matched the choice proportions observed in the experiment. Similar to in Experiment 1, the AOD model reproduced the choice proportions of both choice sets better than all other models.

In the second step, we aggregated choice proportions into bins of 10 trials to see if the models correctly reflected the dynamic development of choice proportions (Figure 5, middle panels). As with the other comparison techniques, the best model also more closely resembled the development of choice proportions. In this comparison, the superiority of the AOD over the other models is evident. The 95% PI was narrower for most of the options and more of the empirical choice proportions lay within it. The only qualitative misprediction was the reversed preference ordering of options B and C in choice set S_1. Yet, similar to the empirical data, the uncertainty was rather high and the 95% PIs overlapped considerably.

Discussion

In Experiment 2, we explored the compromise effect in a decisions-from-experience setting. We used a bandit task with full feedback where participants chose among three options in two choice sets. Two of the options in each of the choice sets formed extreme options and the third option was in between the other ones. We observed behavior systematically violating IIA. However, contrary to the predictions of the compromise effect in decisions-from-description literature, we observed its reversal: a systematic preference for one of the two extreme options, behavior in line with the similarity mechanism of the AOD model. In a model comparison procedure involving the WAIC and posterior predictive checks, the AOD clearly outperformed all other RL models.

Experiment 3

In the third experiment, we explored the attraction effect in a decisions-from-experience setting. We used the same procedure and paradigm as in Experiments 1 and 2 but selected options in the experimental phase to be similar to those in the decisions-from-description literature that demonstrated the attraction effect. As in Experiment 2, we expected the reversal of an attraction effect to arise: The similarity mechanism predicts inhibition of close options, disregarding any dominance relationship, thus lowering their choice proportions. Consequently, adding an option to the choice that is similar to and dominated by another option will *lower* the choice proportion for this dominating option.

Method

Participants and procedure. Twenty-three participants (14 female, age: 20–43 years, $M = 25.91$ years, $SD = 5.49$) participated in Experiment 3. The options were positioned according to an attraction effect by making the similar decoys from Experiment 1 be dominated by the core options. Perception of dominance is crucial for the attraction effect to arise (cf. Huber et al., 2014, p. 522 f.), we thus reduced the options' standard deviations to stress this aspect. Otherwise, the experimental procedure was identical to Experiments 1 and 2.

Paradigm and design. The paradigm, the training phase, and the post questionnaire were identical to those of Experiments 1 and 2. Only the choice sets in the experimental phase were different. Choice set 1 (S_1) contained options A: ($\mu_A = 73$, $\pi_A = .17$), B: ($\mu_B = 78$, $\pi_B = .22$), and C: ($\mu_C = 22$, $\pi_C = .78$), so that option B dominated option A, as it provided on average a larger payoff with a larger probability. Choice set 2 (S_2) contained options B, C, and D: ($\mu_D = 17$, $\pi_D = .73$), so that option C dominated option D. In contrast to the other experiments, the variance of the options was lower than in the training phase, $\sigma \approx 1.29$. This was done such that the draws from the normal distributions had minimal overlap, stressing the dominance relationship. Again, note that between the two choice sets, options B and C were represented by

different stimuli (so that participants had to relearn the options' values). In contrast to Experiment 1, options A and D were generated by taking B and C, respectively, and reducing the mean of the normal distribution μ by 5 points and the probability of drawing from the normal distribution π by .05. Again, contrasting the choice proportions between the two choice sets allowed us to measure the emergence of context effects. We used the RST, this time with the dominating option as target in each of the choice sets, B in S_1 and C in S_2.

Results

For the training phase, accuracy was assessed as the relative proportion of HV choices. The 95% BCI of this variable, *Acc*, was entirely above .5, BCI = [.68, .87]; $M_\mu = .78$; $SD_\mu = .05$, showing that participants were more likely to choose the HV option than the LV one, thus reflecting an ability to learn.

In the experimental phase, the attraction-effect analysis revealed that the 95% BCI of RST was mostly below but included .5, BCI = [.45, .51]; $M_\mu = .48$; $SD_\mu = .02$. In both choice sets, participants preferred option C over the other options. The strong preference for C in S_1 became weaker in S_2. Thus, instead of boosting the preference for C in S_2, the attraction decoy led to a weaker preference for C over B: The presence of D "hurt" the similar option C more than the dissimilar option B, a pattern similar to that found in Experiment 1 and indicating a *reversed* attraction effect (Figure 4, top right panel).

This conclusion is reflected in the analysis of weak IIA violations. None of the participants showed a preference reversal consistent with an attraction effect (preference for B over C in S_1, and C over B in S_2). Instead, 13% showed a preference reversal in the opposite direction, preferring C over B in choice set S_1 and B over C in choice set S_2. Such a pattern supports the notion of a true violation of IIA in the direction opposite the attraction effect, and not merely noise.

The analysis of the post questionnaire about the options' properties did not support the notion of explicit knowledge. Participants' ranks in the estimations of

probabilities and magnitudes did not correlate with how often they chose the options (all $ps \geq .06$).

Model Comparison

In Experiment 3, the baseline model had a deviance of 12,129, meaning that models with adjusted-for-complexity deviances below that threshold explained the data beyond pure guessing. All of our candidate models surpassed this criterion. The AOD outperformed all the other models, having the by-far lowest WAIC (9,914) compared to the runner-up RPR (WAIC: 10,094; see Table 1 for all information criteria). The ΔWAIC of the two best models was 180 points on the deviance scale and the standard error of the differences was 24.08, resulting in a standardized effect size of $\Delta = 7.48\sigma$, thus reflecting a substantially better fit of the AOD model.

A consistent result across the group-level posterior distributions of all models was the rather slight adaptation to recent outcomes, as reflected in the overall low learning rates. Instead, participants included even early observations in their expectations. Additionally, participants were risk averse. For the two models reflecting risk attitudes in the curvature of the utility function, RPR and AOD, the estimates of γ were mostly below 1. The 95% BCI did not include 1 in either case. In the RSE, participants weighted negative RPEs higher than positives ones, again reflecting risk aversion. As in the other experiments, inhibition played an important role in forming participants' impressions of the options. Similar to in Experiment 2, the 95% BCI of the group-level mean of η was mostly, but not entirely, above 0. See Table 4 for posterior distributions of all the models' group-level parameters.

Posterior Predictive Checks

To obtain an impression of the models' absolute fits to the data we performed posterior predictive checks as described in Experiment 1. In the first step, we checked whether the models could reproduce the empirically observed choice proportions (see Figure 4, right column). The general pattern observed in the model comparison using the WAIC was also reflected in the posterior predictive checks. All of the risk-including

models' predictions closely matched the choice pattern in choice set S_1 and were visibly better than the FLR model. Yet, in this experiment all models seem to have missed important aspects of the data. All models mispredicted the choice proportions of options C and D in choice set S_2. In the data, these choice proportions differed by a large margin (.54 vs. .12), while the models predicted choice proportions of around .45 and .20 for options C and D, respectively.

In the second step, we aggregated choice proportions into bins of 10 trials to see if the models correctly reflected the dynamic development of choice proportions (Figure 5, right panels). As with the other comparison techniques, the AOD model's predictions more closely resembled the development of choice proportions in choice set S_1. Despite this, as already seen in the aggregated results, all models underpredicted choices of C and overpredicted choices of D in choice set S_2. Especially for D, almost none of the empirical data lay within the 95% PI.

Discussion

In Experiment 3, we explored the attraction effect in a decisions-from-experience setting. We used a bandit task with full feedback where participants chose among three options in two choice sets. Two of the options in each of the choice sets were similar to each other, with one of them being dominated by the other, and the third option was dissimilar to both of them. We observed behavior violating IIA, although we did not obtain conclusive evidence. The direction of the effect was opposite to the standard attraction effect: Participants preferred the dissimilar option, again providing evidence in favor of the similarity mechanism. In a model comparison procedure involving the WAIC and posterior predictive checks, the AOD model clearly outperformed all other RL models.

Iowa Gambling Task

The AOD model successfully predicted the observed similarity effect in Experiment 1 and the reversal of the compromise and attraction effects in Experiments 2 and 3, respectively, using an experience-based decision-making task. To also assess

the generalizability of the AOD model, we wanted to test its predictions in a standard learning task such as the IGT. The IGT is one of the most widely used tasks to study individuals' learning, risk-taking behavior, and impulsivity behavior (cf. Bechara, Damasio, Tranel, & Damasio, 1997; Busemeyer & Stout, 2002). We used data of the IGT with full feedback for this purpose, in which a healthy population was compared with a population of drug-abusing individuals (Yechiam, Stout, Busemeyer, Rock, & Finn, 2005) and from which we used the sample of healthy individuals. This sample was also analyzed to compare different models of the IGT (Yechiam & Busemeyer, 2005). The critical question is whether the AOD model is able to accurately describe the learning process and the dynamic development of choices.

Method

Seventy-eight participants with complete data participated in this experiment. They were recruited in bars and on the campus of Indiana University, Bloomington, and completed a computerized variant of the original IGT (Bechara et al., 1994) with full feedback. In this task, participants were initially endowed with \$20. They repeatedly chose from one of four decks of cards for a total of 150 trials. Each deck *always* yielded a deck-specific fixed gain, and sometimes *additionally* a loss. Participants obtained both the gain and potential loss of the chosen option on each trial and saw the outcomes of the other three decks as well. After completing the 150 trials, participants received the money they had left at the end of the task (during the task, they always saw the current tally) in addition to the show-up fee of \$7 per hour. The four options, A, B, C, and D, were characterized by two properties: advantageousness and frequency of losses. Decks A and B were disadvantageous, yielding a net loss of \$2.50 every 10 trials, wheres decks C and D were advantageous, yielding a net gain of \$2.50 every 10 trials. Decks A and C had frequent losses, $p(\text{loss}) = .50$, and decks B and D had rare losses, $p(\text{loss}) = .10$. For the frequent-loss options, some noise was added to the outcomes, and some participants received higher gains and losses (50% higher). See Yechiam et al. (2005) for further details. For simplicity we assumed that observed loss and gain

outcomes in one trial were summed up to a total outcome, which we used for estimating the models' parameters.

Model Comparison

For this experiment, the baseline model predicting each choice with equal probability had a deviance of 32,439, meaning that models with adjusted-for-complexity deviances below that threshold explained the data beyond pure guessing. All of our candidate models surpassed this criterion. The AOD model outperformed all the other models, having the by-far lowest WAIC (24,823) compared to the runner-up RSE (WAIC: 27,056; see Table 1 for all information criteria). The ΔWAIC of the two best models was 2,233 points on the deviance scale and the standard error of the differences was 94.09, resulting in a standardized effect size of $\Delta = 23.73\sigma$, reflecting a substantially better fit of the AOD model.

A consistent result across the group-level posterior distributions of all models was the rather slight adaptation to recent outcomes, as reflected in the overall low learning rates. Instead, participants included even early observations in the formation of their expectations. In terms of risk preferences, the models disagreed. For the two models reflecting risk attitudes in the curvature of the utility function, RPR and AOD, estimates of γ were close to 0. In the model with separate learning rates for positive and negative RPEs, RSE, participants weighted positive RPEs higher than negative ones, reflecting risk seeking. As in the previous experiments, the similarity mechanism played an important role in forming participants' impressions: The 95% BCI of the group-level mean of η was mostly, albeit not entirely, above 0. This indicates that options that yielded similar outcomes were inhibited and thus perceived as less attractive than options with distinct outcomes. See Table 5 for posterior distributions of all the models' group-level parameters.

Posterior Predictive Checks

We performed posterior predictive checks not only to evaluate the model fits relative to each other but also to obtain an impression of the models' absolute fits to

the data. We used the same procedure as in all other experiments but performed only the more informative analysis of choice-proportion development. We aggregated choice proportions into bins of 10 trials to see if the models correctly reflected the dynamic development of choice proportions (Figure 6). Similar to the other model comparison criteria, the dynamic development of the AOD model also most closely resembled the development of choice proportions. It captured the quickly developing and afterward stable preference for B, the slowly increasing preference for D, and the slowly decreasing preference for the remaining options. The other two rather similarly performing models, RPR and RSE, predicted a slower development of preference for B. The RPR predicted a clear preference for C over A, a pattern not present in the data, whereas the RSE model evidently mispredicted the development of the choice proportions of D. The 95% PI of the AOD model clearly included many more empirical data than all of the other models.

Discussion

In addition to our own experiments designed to test context effects, we also examined how well the AOD model was able to describe learning processes in a standard learning task such as the IGT. The IGT is a paradigm not specifically designed to examine similarity effects. Again the AOD model, the only model predicting IIA violations, clearly won the model comparison. Participants' behavior seems to have violated IIA in this task as well. In line with the parameter estimates in the previous experiments, similar options were perceived as less attractive than dissimilar ones.

General Discussion

The present study systematically explored violations of IIA in decision-from-experience tasks. We rigorously tested a novel, psychologically motivated model of RL that predicts violations of IIA on the basis of a similarity mechanism. In a series of three experiments, we tested whether the similarity effect, the compromise effect, and the attraction effect also occurred for a decision-from-experience task with full feedback. We observed a similarity effect that represented a violation of strong IIA.

However, we observed neither the attraction nor the compromise effect. Instead, we observed a reversed compromise effect and a weak reversed attraction effect, both violating strong IIA. The observed violations of weak IIA support our conclusions from the quantitative analyses. Additionally, participants lacked explicit knowledge of the underlying payoff distributions. The behavioral findings were reflected in the model comparison of the cognitive models. The suggested AOD model outperformed all alternative learning models in predicting the observed behavior and the violations of IIA. The model comparison in the IGT supported the conclusions drawn from our own experiments.

Our behavioral findings provide evidence that IIA is violated not only in decisions from description but also in decisions from experience, more specifically in decisions based on learning. Most strikingly, the violations of IIA that we observed were mostly *not* in line with traditional context-effect research. While we elicited behavior consistent with the predictions of the similarity effect, we observed neither the compromise effect nor the attraction effect. Instead, we observed reversals of these two effects. These reversals of both the compromise and the attraction effect were predicted by the similarity mechanism of the AOD model.

One notable behavioral finding is that the reversal of the attraction effect in Experiment 3 was comparatively weak. The slight shift in the outcome distributions of the decoy options as compared to Experiment 1 resulted in a much weaker effect, leading to the inclusion of .5 in the BCI of the RST. Defending the attribute-based explanation of the attraction effect, one could argue that this shift resulted from the more apparent dominance relationship, as the outcomes were not only less frequent, but also consistently smaller than those of the target option. Furthermore, all of the standard deviations were reduced, thus stressing this relationship. Despite its compatibility with traditional interpretations of context effects, this argumentation lacks empirical evidence, as participants did not show any explicit knowledge about this relationship. Moreover, the cognitive modeling results corroborate the similarity mechanism.

Model Comparison

In the present work, we introduced a new model that incorporates a similarity mechanism, the AOD model. The AOD model follows traditional learning models by assuming a standard updating-learning mechanism. It represents different risk attitudes by using a utility function that allows for risk-averse and risk-seeking behavior. Contrary to standard learning models, the AOD model additionally assumes that choice options inhibit each other as a function of their similarity so that similar options appear relatively less attractive than dissimilar options. We rigorously compared the AOD model with three other RL models (FLR, RPR, and RSE) and found that it systematically outperformed all of them by a wide margin. This finding is most remarkable in the case of Experiment 3. As mentioned in the previous paragraph, it is the only one of the three experiments not showing a systematic (behavioral) violation of IIA. Nonetheless, the AOD performed better than models that assume that the subjective value of an option is independent of the other options (i.e., the updating process of the subjective expectations is performed independently). As such, inhibition-based competition of options performs better than no inhibition at all—a notion clearly refuting traditional accounts of the context effects, as they would predict the attraction effect. Furthermore, the FLR model and the RPR model (but not the RSE model) are nested within the AOD model. Since the RSE model performed rather poorly anyway, this allows us to infer that the additional mechanism assumed by the AOD model is responsible for its superior performance. To stress the generality of the AOD model for experience-based decisions, we have also shown that it is possible to apply the AOD model to compatible paradigms, such as the IGT with full feedback (Yechiam et al., 2005). The IGT was not designed to test context effects. Nevertheless, the AOD model with its similarity mechanism still provided a better description of the observed learning processes in this task.

Full Versus Partial Feedback

It is important to stress that in the present study we looked at violations of IIA only in experienced-based decision-making tasks with full feedback. In contrast to full feedback, partial feedback only provides feedback about the chosen option. In partial-feedback situations, the nonchosen options have to be chosen in subsequent trials to obtain information about their underlying outcome distributions. This introduces the so-called exploration–exploitation dilemma (e.g., Auer, 2003; Cohen, McClure, & Yu, 2007; Frank et al., 2009; Navarro, Newell, & Schulze, 2016). Exploration is defined as choosing options other than the one with the highest subjective expectation to obtain more information about the outcome distribution. Exploitation, that is, choosing the option with the highest subjective expectation and thus maximizing the subjective reward, is the direct opposite of exploration. Full feedback, and forgone payoffs that come with it, render exploration behavior irrelevant, as the information about the forgone payoffs comes free of charge. Although it is sometimes argued that humans might pay less attention to forgone payoffs (Yechiam & Busemeyer, 2005), models proposing a separate learning of forgone payoffs are at risk of overfitting mere choice patterns, irrespective of the observed outcomes (cf. Yechiam & Ert, 2007). Nevertheless, forgone payoffs heavily influence participants' behavior: With full feedback, participants maximize expected value more often (Rakow, Newell, & Wright, 2015), tend to behave with less risk aversion (Yechiam & Busemeyer, 2006), and show different behavioral phenomena than in partial-feedback conditions (Plonsky & Erev, 2017; Plonsky et al., 2015). With partial feedback, it is difficult to derive clear predictions about a potential similarity effect, and future research would have to clarify this condition. Nevertheless, we defined the AOD model in a way that it can be applied to partial-feedback situations. Having less information about the options—in the case of three options, participants have a third of the information available—might lead to more exploration (i.e., more random behavior), thus mitigating the similarity effect. Additionally, comparing outcomes of the options with each other becomes more difficult, as the unchosen outcomes are not shown in each trial and would have to be memorized.

Interindividual Differences

In the present study, the proposed similarity mechanism is mainly reflected in the η parameter of the AOD model, which governs the balance between subjective utility and similarity-based inhibition. The posterior distribution of the group-level mean was mostly above 0 in all experiments. Moreover, Experiments 2, 3, and the IGT, the 95% BCI included 0, showing that on the group level there is no clear evidence for the similarity mechanism in one direction. Together with the evidently better fit of the AOD model compared to the other models, these results reflect high interindividual variability in the way similarity between choice outcomes is integrated into the decision process. This variability becomes evident in the behavioral data by grouping participants who show similarity-based inhibition (i.e., their posterior distributions of the individual-level η parameters do not include 0) and contrasting them with participants who do not show similarity-based inhibition (i.e., whose posterior distributions of η overlap with 0). Figure 7 illustrates the choice proportions in the experiments after conducting this split. For Experiments 1–3 (aggregated across experiments, top row of Figure 7), there is a visible difference between these two subgroups: Only participants who are best described by using the similarity mechanism show clear violations of IIA, whereas this pattern is much weaker for participants for whom the similarity mechanism is not essential. In the IGT (bottom row of Figure 7), there are no behavioral patterns that clearly indicate violations of IIA. Nevertheless, there are behavioral differences between the participants best described by the similarity mechanism and those participants for whom the mechanism is not essential. The most prominent difference is that participants described by the similarity mechanism have an even stronger preference for the disadvantageous, rare-loss deck B (at the expense of the other disadvantageous-but-frequent-loss deck A). Between the two advantageous decks there are only minor differences. These findings fit into a larger body of decisions-from-description literature showing that interindividual differences play an important role in which context effects arise, and when (Liew, Howe, & Little, 2016; Trueblood et al., 2015).

Future Directions

In the future, it will be important to further explore the similarty mechanism and the generalizability of the AOD model. How do context-dependent preferences manifest themselves in other constellations, for example, when there are outcome distributions with more than two modes? How do dynamic outcome distributions (i.e., outcome distributions that change with time) influence choice behavior (e.g., Nassar, Wilson, Heasly, & Gold, 2010)? What influence do correlated/anticorrelated outcomes exert (e.g., Experiment 4 in Tsetsos, Chater, & Usher, 2012)? What factors drive the interindividual differences?

The questions of correlated/anticorrelated outcomes and of the AOD model's generalizability are closely related to each other, as the model makes strong predictions due to its functional form: Besides the outcomes' absolute magnitudes, similarity-based inhibition of options is based on the correlation between the outcomes, and correlated or anticorrelated outcomes heavily influence the model's predictions.

Conclusion

To our knowledge, the present study is the first to demonstrate systematic violations of IIA in decisions from experience. The violations we observed, however, are mostly not in line with traditional context-effect research. The results contribute first experimental evidence for what might become another description–experience gap. Additionally, our work illustrates the shortcomings of standard context-insensitive RL models and provides a starting point for incorporating psychological insights into the field of RL. Future discoveries of new phenomena in decisions from experience and the development of other psychologically motivated models will improve the understanding of the cognitive processes underlying context sensitivity in human learning and decision making.

References

Auer, P. (2003). Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, *3*, 397–422.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7–15. doi:10.1016/0010-0277(94)90018-3

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*, 1293–1295. doi:10.1126/science.275.5304.1293

Becker, G. M., Degroot, M. H., & Marschak, J. (1963). Probabilities of choices among very similar objects: An experiment to decide between two models. *Behavioral Science*, *8*, 306–311. doi:10.1002/bs.3830080403

Berkowitsch, N. A. J., Scheibehenne, B., & Rieskamp, J. (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General*, *143*, 1331–1348. doi:10.1037/a0035159

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, *120*, 522–543. doi:10.1037/a0032457

Brevers, D., Bechara, A., Cleeremans, A., & Noël, X. (2013). Iowa gambling task (IGT): Twenty years after – gambling disorder and IGT. *Frontiers in Psychology*, *4*, 1–14. doi:10.3389/fpsyg.2013.00665

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment*, *14*, 253–262. doi:10.1037/1040-3590.14.3.253

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*, 1–32. doi:10.18637/jss.v076.i01

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration.

Philosophical Transactions of the Royal Society B: Biological Sciences, *362*, 933–942. doi:10.1098/rstb.2007.2098

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*, 1204–1215. doi:10.1016/j.neuron.2011.02.027

Doyle, J. R., O'Connor, D. J., Reynolds, G. M., & Bottomley, P. A. (1999). The robustness of the asymmetrically dominated effect: Buying frames, phantom alternatives, and in-store purchases. *Psychology and Marketing*, *16*, 225–243. doi:10.1002/(SICI)1520-6793(199905)16:3<225::AID-MAR3>3.0.CO;2-X

Erev, I., & Haruvy, E. (2005). Generality, repetition, and the role of descriptive learning models. *Journal of Mathematical Psychology*, *49*, 357–371. doi:10.1016/j.jmp.2005.06.009

Erev, I., & Roth, A. E. (2014). Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences*, *111*, 10818–10825. doi:10.1073/pnas.1402846111

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience*, *12*, 1062–1068. doi:10.1038/nn.2342

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC Press.

Gershman, S. J. (2015). Do learning rates adapt to the distribution of rewards? *Psychonomic Bulletin & Review*, *22*, 1320–1327. doi:10.3758/s13423-014-0790-3

Gluth, S., Hotaling, J. M., & Rieskamp, J. (2017). The attraction effect modulates reward prediction errors and intertemporal choices. *Journal of Neuroscience*, *37*, 371–382. doi:10.1523/JNEUROSCI.2532-16.2017

Heath, T. B., & Chatterjee, S. (1995). Asymmetric decoy effects on lower-quality versus higher-quality brands: Meta-analytic and experimental evidence. *Journal of Consumer Research*, *22*, 268–284. doi:10.1086/209449

Herne, K. (1999). The effects of decoy gambles on individual choice. *Experimental Economics*, *2*, 31–40. doi:10.1023/A:1009925731240

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*, 534–539. doi:10.1111/j.0956-7976.2004.00715.x

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, *13*, 517–523. doi:10.1016/j.tics.2009.09.004

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623. arXiv: 1111.4246

Hotaling, J. M., Busemeyer, J. R., & Li, J. (2010). Theoretical developments in decision field theory: Comment on Tsetsos, Usher, and Chater (2010). *Psychological Review*, *117*, 1294–1298. doi:10.1037/a0020401

Hu, J., & Yu, R. (2014). The neural correlates of the decoy effect in decisions. *Frontiers in Behavioral Neuroscience*, *8*, 1–8. doi:10.3389/fnbeh.2014.00271

Huber, J., Payne, J. W., & Puto, C. P. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*, 90–98. doi:10.1086/208899

Huber, J., Payne, J. W., & Puto, C. P. (2014). Let's be honest about the attraction effect. *Journal of Marketing Research*, *51*, 520–525. doi:10.1509/jmr.14.0208

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–292. doi:10.2307/1914185

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*, 1292–1298. doi:10.1038/nn.2635

Krause, F., & Lindemann, O. (2014). Expyriment: A Python library for cognitive and neuroscientific experiments. *Behavior Research Methods*, *46*, 416–428. doi:10.3758/s13428-013-0390-6

Liew, S. X., Howe, P. D. L., & Little, D. R. (2016). The appropriacy of averaging in the study of context effects. *Psychonomic Bulletin & Review*, *23*, 1639–1646. doi:10.3758/s13423-016-1032-7

Link, W. A., & Eaton, M. J. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, *3*, 112–115. doi:10.1111/j.2041-210X.2011.00131.x

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York, NY: Wiley.

Mikhael, J. G., & Bogacz, R. (2016). Learning reward uncertainty in the basal ganglia. *PLOS Computational Biology*, *12*, 1–28. doi:10.1371/journal.pcbi.1005062

Mochon, D. (2013). Single-option aversion. *Journal of Consumer Research*, *40*, 555–566. doi:10.1086/671343

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, *30*, 12366–12378. doi:10.1523/JNEUROSCI.0822-10.2010

Navarro, D. J., Newell, B. R., & Schulze, C. (2016). Learning and choosing in an uncertain world: An investigation of the explore–exploit dilemma in static and dynamic environments. *Cognitive Psychology*, *85*, 43–77. doi:10.1016/j.cogpsych.2016.01.001

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience*, *32*, 551–562. doi:10.1523/JNEUROSCI.5498-10.2012

Noguchi, T., & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition*, *132*, 44–56. doi:10.1016/j.cognition.2014.03.006

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*, 375–402.

Parrish, A. E., Evans, T. A., & Beran, M. J. (2015). Rhesus macaques (Macaca mulatta) exhibit the decoy effect in a perceptual discrimination task. *Attention, Perception, & Psychophysics*, *77*, 1715–1725. doi:10.3758/s13414-015-0885-6

Pettibone, J. C. (2012). Testing the effect of time pressure on asymmetric dominance and compromise decoys in choice. *Judgment and Decision Making*, *7*, 513–521. Retrieved from http://journal.sjdm.org/11/111114/jdm111114.pdf

Pettibone, J. C., & Wedell, D. H. (2000). Examining models of nondominated decoy effects across judgment and choice. *Organizational Behavior and Human Decision Processes*, *81*, 300–328. doi:10.1006/obhd.1999.2880

Plonsky, O., & Erev, I. (2017). Learning in settings with partial feedback and the wavy recency effect of rare events. *Cognitive Psychology*, *93*, 18–43. doi:10.1016/j.cogpsych.2017.01.002

Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review*, *122*, 621–647. doi:10.1037/a0039413

Rakow, T., Newell, B. R., & Wright, L. (2015). Forgone but not forgotten: The effects of partial and full feedback in "harsh" and "kind" environments. *Psychonomic Bulletin & Review*, *22*, 1807–1813. doi:10.3758/s13423-015-0848-x

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, *44*, 631–661. doi:10.1257/jel.44.3.631

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review*, *108*, 370–392. doi:10.1037/0033-295X.108.2.370

Scarpi, D. (2011). The impact of phantom decoys on choices in cats. *Animal Cognition*, *14*, 127–136. doi:10.1007/s10071-010-0350-9

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599. doi:10.1126/science.275.5306.1593

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*, 158–174. doi:10.1086/209205

Soltani, A., De Martino, B., & Camerer, C. (2012). A range-normalization model of context-dependent choice: A new model and evidence. *PLoS Computational Biology*, *8*, 1–15. doi:10.1371/journal.pcbi.1002607

Stan Development Team. (2016). PyStan: The Python interface to Stan. Retrieved from http://mc-stan.org

Steingroever, H., Wetzels, R., & Wagenmakers, E.-J. (2013). A comparison of reinforcement learning models for the Iowa gambling task using parameter space partitioning. *Journal of Problem Solving*, *5*, 1–32. doi:10.7771/1932-6246.1150

Steingroever, H., Wetzels, R., & Wagenmakers, E.-J. (2014). Absolute performance of reinforcement-learning models for the Iowa gambling task. *Decision*, *1*, 161–183. doi:10.1037/dec0000005

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tan, K., Dong, S., Liu, X., Chen, W., Wang, Y., Oldroyd, B. P., & Latty, T. (2015). Phantom alternatives influence food preferences in the eastern honeybee Apis cerana. *Journal of Animal Ecology*, *84*, 509–517. doi:10.1111/1365-2656.12288

Trueblood, J. S. (2012). Multialternative context effects obtained using an inference task. *Psychonomic Bulletin & Review*, *19*, 962–968. doi:10.3758/s13423-012-0288-9

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, *121*, 179–205. doi:10.1037/a0036137

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2015). The fragile nature of contextual preference reversals: Reply to Tsetsos, Chater, and Usher (2015). *Psychological Review*, *122*, 848–853. doi:10.1037/a0039656

Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers: Context effects are fundamental to decision making. *Psychological Science*, *24*, 901–908. doi:10.1177/0956797612464241

Trueblood, J. S., & Pettibone, J. C. (2017). The phantom decoy effect in perceptual decision making. *Journal of Behavioral Decision Making*, *30*, 157–167. doi:10.1002/bdm.1930

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, *109*, 9659–9664. doi:10.1073/pnas.1119569109

Tsetsos, K., Chater, N., & Usher, M. (2015). Examining the mechanisms underlying contextual preference reversal: Comment on Trueblood, Brown, and Heathcote (2014). *Psychological Review*, *122*, 838–847. doi:10.1037/a0038953

Tsetsos, K., Usher, M., & Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological Review*, *117*, 1275–1293. doi:10.1037/a0020580

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299. doi:10.1037/h0032955

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323. doi:10.1007/BF00122574

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*, 757–769. doi:10.1037/0033-295X.111.3.757

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432. doi:10.1007/s11222-016-9696-4

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594. arXiv: 1004.2316

Wedell, D. H. (1991). Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 767–778. doi:10.1037//0278-7393.17.4.767

Wetzels, R., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2010). Bayesian parameter estimation in the expectancy valence model of the Iowa gambling task. *Journal of Mathematical Psychology*, *54*, 14–27. doi:10.1016/j.jmp.2008.12.001

Wollschläger, L. M., & Diederich, A. (2012). The 2N-ary choice tree model for N-alternative preferential choice. *Frontiers in Psychology*, *3*, 1–11. doi:10.3389/fpsyg.2012.00189

Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review*, *12*, 387–402. doi:10.3758/BF03193783

Yechiam, E., & Busemeyer, J. R. (2006). The effect of foregone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making*, *19*, 1–16. doi:10.1002/bdm.509

Yechiam, E., & Busemeyer, J. R. (2008). Evaluating generalizability and parameter consistency in learning models. *Games and Economic Behavior*, *63*, 370–394. doi:10.1016/j.geb.2007.08.011

Yechiam, E., & Ert, E. (2007). Evaluating the reliance on past choices in adaptive learning models. *Journal of Mathematical Psychology*, *51*, 75–84. doi:10.1016/j.jmp.2006.11.002

Yechiam, E., Stout, J. C., Busemeyer, J. R., Rock, S. L., & Finn, P. R. (2005). Individual differences in the response to forgone payoffs: An examination of high functioning drug abusers. *Journal of Behavioral Decision Making*, *18*, 97–110. doi:10.1002/bdm.487

Zeigenfuse, M. D., Pleskac, T. J., & Liu, T. (2014). Rapid decisions from experience. *Cognition*, *131*, 181–194. doi:10.1016/j.cognition.2013.12.012

Table 1
Information Criteria for Each of the Models in All Experiments

	Model	$\widehat{\text{elpd}}_{\text{WAIC}}$	p_{WAIC}	WAIC	SE_{WAIC}
Experiment 1	FLR	12,103	22	12,146	38.15
	RPR	10,868	53	10,973	75.38
	AOD	10,573	70	10,712	78.57
	RSE	11,079	57	11,192	71.81
Experiment 2	FLR	10,888	34	10,956	57.57
	RPR	10,412	51	10,514	68.56
	AOD	9,938	67	10,072	76.39
	RSE	10,421	57	10,535	68.58
Experiment 3	FLR	11,051	37	11,126	53.13
	RPR	9,964	65	10,094	88.4
	AOD	9,752	81	9,914	87.51
	RSE	10,052	69	10,189	87.24
Iowa gambling task	FLR	29,730	190	30,109	109.38
	RPR	26,703	217	27,138	131.03
	AOD	24,147	338	24,823	146.33
	RSE	26,696	180	27,056	129.93

Note. FLR = Fixed-learning-rate reinforcement learning model. RPR = Risk-preference reinforcement learning model. AOD = Accentuation of differences model. RSE = Risk-sensitive reinforcement learning model. $\widehat{\text{elpd}}_{\text{WAIC}}$ = expected sum of log-pointwise predictive deviances. p_{WAIC} = effective number of parameters. WAIC = widely applicable information criterion (Watanabe, 2010). SE_{WAIC} = standard error of the WAIC. All measures are reported on the deviance scale and for the complete data sets. For log pointwise predictive densities, all numbers have to be divided by $-2 \times 2,880$ and $-2 \times 11,700$ for Experiment 1–3 and the Iowa gambling task, respectively.

Table 2
Posterior Distributions of the Models' Group-Level Mean Parameters in Experiment 1

		Posterior percentile					
Model	Parameter	2.5%	25%	50%	75%	97.5%	μ
FLR	θ	0.13	0.26	0.39	0.58	1.25	0.39
	α	.00	.01	.03	.06	.19	.03
RPR	θ	0.63	0.92	1.10	1.34	2.07	1.11
	α	.01	.03	.05	.08	.19	.05
	γ	0.29	0.41	0.48	0.56	0.71	0.48
AOD	θ	0.75	1.11	1.33	1.62	2.44	1.34
	α	.01	.03	.06	.09	.21	.06
	γ	0.19	0.29	0.35	0.42	0.57	0.34
	ψ	0.17	0.38	0.58	0.90	2.31	0.59
	η	0.69	1.55	1.97	2.4	3.19	1.97
RSE	θ	0.27	0.42	0.54	0.71	1.23	0.55
	α^+	.00	.01	.01	.01	.03	.01
	α^-	.02	.06	.09	.14	.28	.09

Note. FLR = Fixed-learning-rate reinforcement learning model. RPR = Risk-preference reinforcement learning model. AOD = Accentuation of differences model. RSE = Risk-sensitive reinforcement learning model.

Table 3
Posterior Distributions of the Models' Group-Level Mean Parameters in Experiment 2

		Posterior percentile					
Model	Parameter	2.5%	25%	50%	75%	97.5%	μ
FLR	θ	0.11	0.23	0.34	0.51	1.21	0.34
	α	.01	.02	.04	.08	.22	.04
RPR	θ	0.06	0.12	0.16	0.23	0.50	0.17
	α	.01	.03	.05	.08	.24	.05
	γ	0.86	1.03	1.12	1.21	1.41	1.12
AOD	θ	0.09	0.16	0.22	0.32	0.67	0.23
	α	.00	.02	.04	.07	.21	.04
	γ	0.82	1.00	1.10	1.19	1.39	1.09
	ψ	0.04	0.11	0.21	0.40	1.31	0.21
	η	-0.36	0.77	1.33	1.89	2.95	1.33
RSE	θ	0.07	0.10	0.13	0.16	0.26	0.13
	α^+	.03	.06	.10	.14	.29	.10
	α^-	.01	.03	.05	.09	.25	.05

Note. FLR = Fixed-learning-rate reinforcement learning model. RPR = Risk-preference reinforcement learning model. AOD = Accentuation of differences model. RSE = Risk-sensitive reinforcement learning model.

Table 4
Posterior Distributions of the Models' Group-Level Mean Parameters in Experiment 3

		Posterior percentile					
Model	Parameter	2.5%	25%	50%	75%	97.5%	μ
FLR	θ	0.29	0.52	0.69	0.92	1.80	0.69
	α	.00	.01	.01	.01	.04	.01
RPR	θ	0.42	0.58	0.67	0.78	1.06	0.67
	α	.01	.01	.02	.03	.06	.02
	γ	0.68	0.78	0.82	0.86	0.94	0.81
AOD	θ	0.42	0.61	0.75	0.93	1.45	0.76
	α	.01	.01	.02	.03	.06	.02
	γ	0.60	0.71	0.76	0.81	0.91	0.76
	ψ	0.08	0.26	0.44	0.79	2.39	0.44
	η	-0.38	0.28	0.62	0.97	1.70	0.63
RSE	θ	0.28	0.38	0.45	0.54	0.84	0.46
	α^+	.01	.01	.01	.02	.03	.01
	α^-	.02	.04	.05	.07	.13	.05

Note. FLR = Fixed-learning-rate reinforcement learning model. RPR = Risk-preference reinforcement learning model. AOD = Accentuation of differences model. RSE = Risk-sensitive reinforcement learning model.

Table 5
Posterior Distributions of the Models' Group-Level Mean Parameters in the Iowa Gambling Task

		Posterior percentile					
Model	Parameter	2.5%	25%	50%	75%	97.5%	μ
FLR	θ	0.40	0.66	0.87	1.15	1.91	0.87
	α	.04	.13	.21	.34	.57	.22
RPR	θ	2.31	2.99	3.41	3.94	5.09	3.43
	α	.05	.08	.10	.13	.21	.10
	γ	0.08	0.11	0.14	0.16	0.22	0.13
AOD	θ	2.35	2.91	3.27	3.70	4.63	3.27
	α	.04	.07	.08	.10	.15	.08
	γ	0.22	0.27	0.30	0.33	0.39	0.29
	ψ	0.25	0.38	0.49	0.62	1.02	0.49
	η	-1.49	-0.37	0.16	0.73	1.84	0.18
RSE	θ	5.09	6.40	7.23	8.22	10.47	7.26
	α^+	.01	.01	.01	.01	.02	.01
	α^-	.00	.00	.00	.00	.00	.00

Note. FLR = Fixed-learning-rate reinforcement learning model. RPR = Risk-preference reinforcement learning model. AOD = Accentuation of differences model. RSE = Risk-sensitive reinforcement learning model.

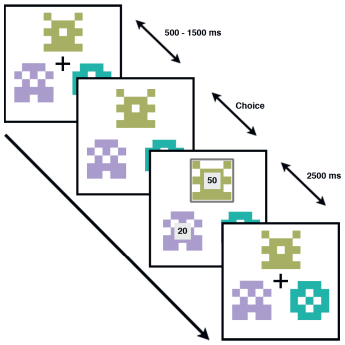


Figure 1. Depiction of a choice trial.

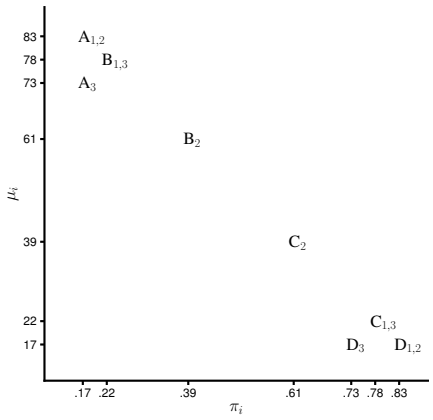


Figure 2. Option placement in the two-dimensional attribute space. Options are characterized by a mean magnitude μ_i and the probability π_i of a draw from a normal distribution with that mean. Subscripts depict experiment in which the option placement was used. For instance, $B_{1,3}$ was used in Experiments 1 and 3 and had $\pi_B = .22$ and $\mu_B = 78$.

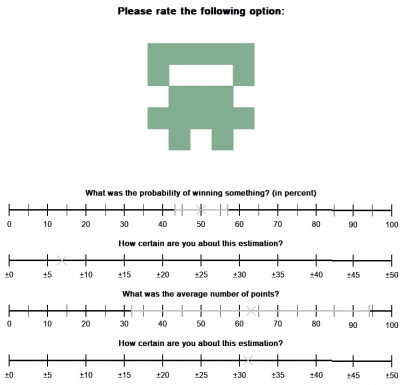


Figure 3. Explicit judgments about options' attributes.

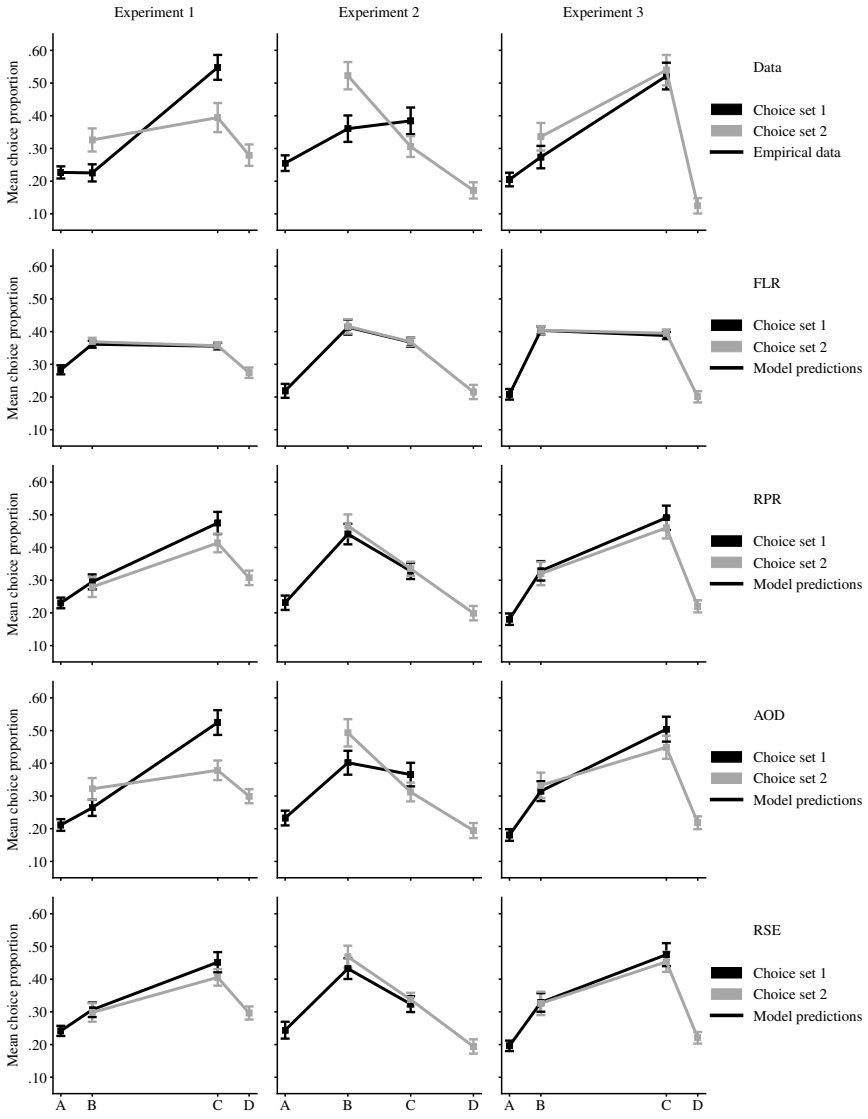


Figure 4. Mean choice proportions of the options in the experiments and choices predicted from the model simulations. Choice set 1 contained options {A, B, C} and choice set 2 contained options {B, C, D}. Options were characterized by a mean magnitude μ_i and the probability π_i of a draw from a normal distribution with that mean. Options A and D served as decoys for options C and B, respectively, in Experiment 1, and vice versa in the other experiments. For illustrative purposes, the distances between the options on the x axis reflect closeness. Options further to the left are riskier and further to the right safer. Error bars indicate ± 1 *SE* and mean *SE* across all simulations for the experiments and simulations, respectively. FLR = Fixed-learning-rate reinforcement learning model. RPR = Risk-preference reinforcement learning model. AOD = Accentuation of differences model. RSE = Risk-sensitive reinforcement learning model.

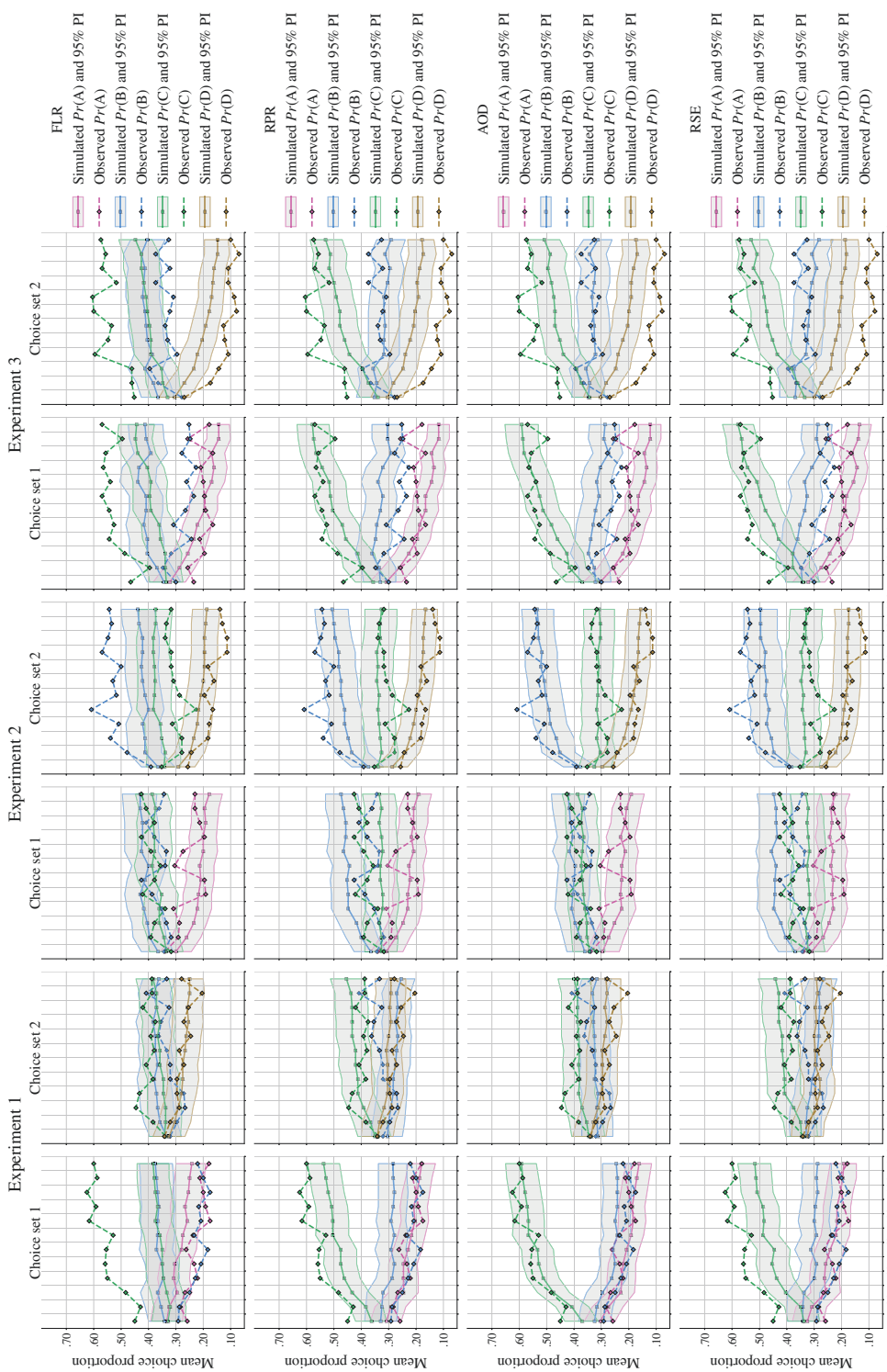


Figure 5. Development of model predictions for each of the options in each of the experiments. Choice set 1 contains options {A, B, C}, choice set 2 contains options {B, C, D}. Options A and D served as decoys for options C and D, respectively, in Experiment 1, and vice versa in the other experiments. FLR = Fixed-learning-rate reinforcement learning model. RPR = Risk-preference reinforcement learning model. AOD = Accentuation of differences model. RSE = Risk-sensitive reinforcement learning model. All points depict mean choice proportions in bins of 10 trials across the simulations. Error bars indicate the 95% Bayesian posterior interval of the mean (PI).

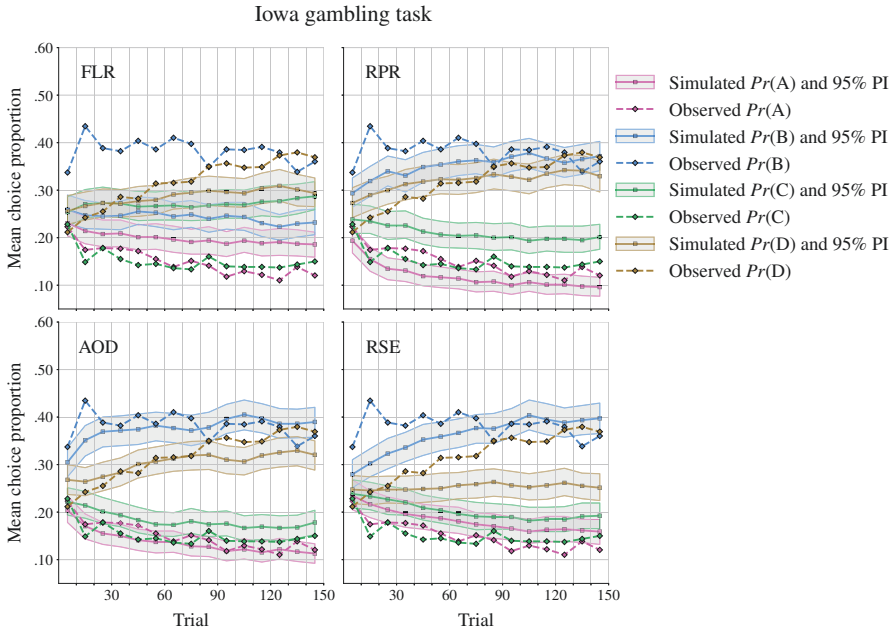


Figure 6. Development of model predictions for each of the options in the Iowa gambling task. FLR = Fixed-learning-rate reinforcement learning model. RPR = Risk-preference reinforcement learning model. AOD = Accentuation of differences model. RSE = Risk-sensitive reinforcement learning model. All points depict mean choice proportions in bins of 10 trials across the simulations. Error bars indicate the 95% Bayesian posterior interval of the mean (PI).

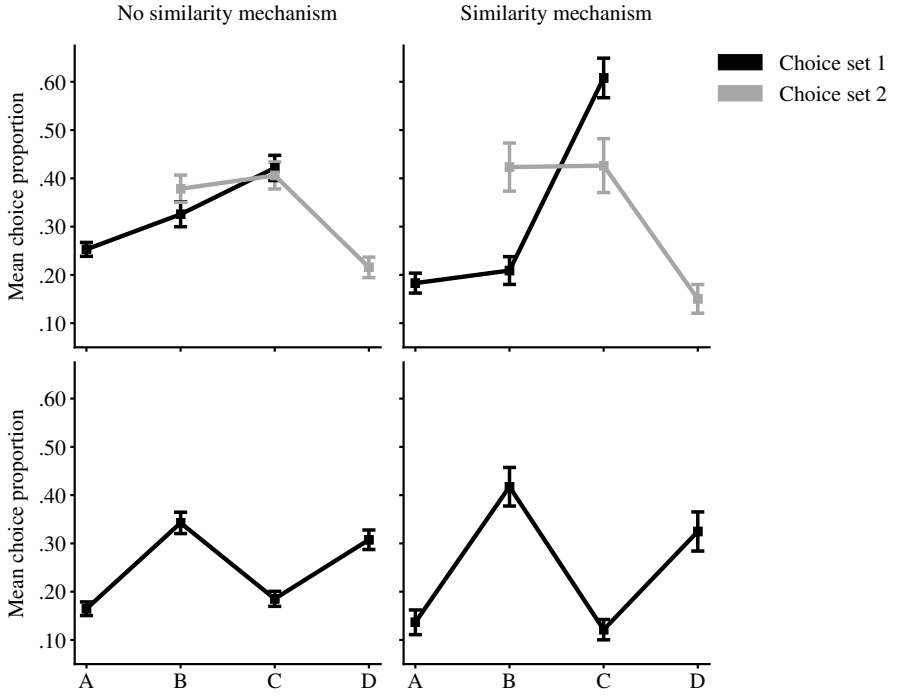


Figure 7. Mean choice proportions of the options in the experiments split by presence of the similarity mechanism. Participants were assigned to the "similarity mechanism" group (right column) if their individual-level posterior distributions of η (accentuation of differences model) were entirely above or below 0. Participants whose posterior distributions overlapped with 0 were assigned to the "no similarity mechanism" group. Top row: Choice proportions aggregated across Experiments 1–3. Choice set 1 contains options {A, B, C}; choice set 2 contains options {B, C, D}. See the respective Method sections for details about the options' properties. Bottom row: Choice proportions in the Iowa gambling task. Error bars indicate ± 1 *SE*.

Appendix

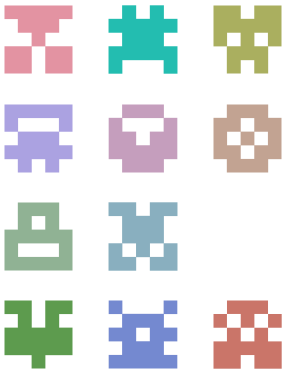


Figure A1. Stimuli used in the experiment. Note that for conciseness, the colors of the stimuli as well as the stimuli themselves represent each of the possibilities and were randomized for each participant. Also note that the colors in the bottom row along with three randomly selected stimuli have only been used in the example during the instructions and not in the experiment itself.

Sampler Settings

All sampling was done in `Stan` 2.14.0 or 2.15.0 (Carpenter et al., 2017) via the `PyStan` interface (Stan Development Team, 2016) using the no-U-turn sampler (Hoffman & Gelman, 2014). We ran four randomly initialized chains in parallel for at least 10,000 total iterations. We used warm-up periods of at least 1,000 iterations until the sampler's parameters were properly tuned, as reflected in each of the parameters' chains (chains resembling "fat hairy caterpillars," as they are commonly called). These warm-up samples were discarded afterward. Of the remaining iterations, we kept at least 1,000 (500 for the IGT) samples per chain (thinning), resulting in a total of at least 4,000 (2,000 for the IGT) samples.[3] Afterward we assessed model convergence by computing $\hat{R}$ (Gelman et al., 2013, p. 285) for each parameter separately across the four chains. If not all $\hat{R} < 1.01$, we restarted the sampling procedure and increased the number of total iterations. In the case of the AOD model fitted to the Iowa gambling task, we could not reach the desired convergence criterion. Nonconvergence leads to an underestimation of model fit and an overpenalization of model complexity. Since we used a very strict convergence criterion and our model outperformed all other models in formal model comparison, these issues of nonconvergence are negligible.

Frequentist Analyses

Experiment 1. Accuracy was assessed as the relative proportion of high-value picks in the training phase. A two-tailed one-sample t test showed a significantly higher probability of selecting HV than LV options, $t(23) = 4.37$; $p < .01$; $d_z = 0.89$. In the experimental phase, the similarity effect was measured as described in the Method section. A two-tailed one-sample t test showed a significant similarity effect, $t(23) = 3.70$; $p < .01$; $d_z = 0.76$.

[3] Although thinning is inefficient (Link & Eaton, 2012), we used it to reduce the amount of data saved.

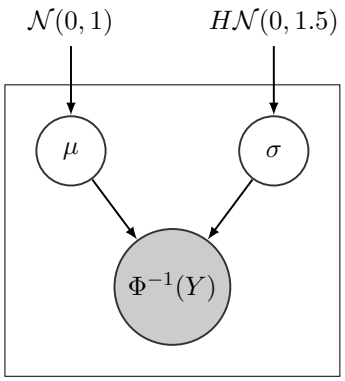


Figure A2. Graphical depiction of the basic Bayesian model. Shaded node represents observed data, unshaded nodes represent model parameters, and distributions without nodes represent priors. $\mathcal{N}$ = Normal distribution; $H\mathcal{N}$ = half-normal distribution.

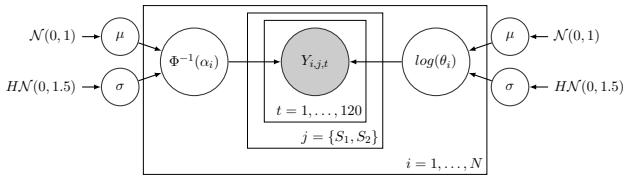


Figure A3. Graphical depiction of the fixed-learning-rate reinforcement learning model. Shaded node represents observed data, unshaded nodes represent model parameters, and distributions without nodes represent priors. $\mathcal{N}$ = Normal distribution; $H\mathcal{N}$ = half-normal distribution.

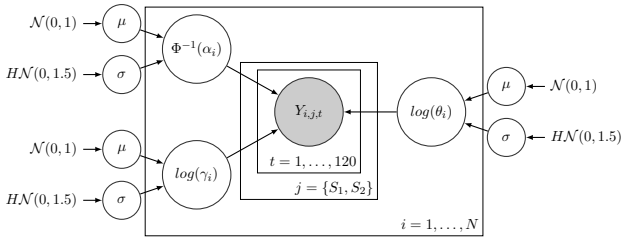


Figure A4. Graphical depiction of the risk-preference reinforcement learning model. Shaded node represents observed data, unshaded nodes represent model parameters, and distributions without nodes represent priors. $\mathcal{N}$ = Normal distribution; $H\mathcal{N}$ = half-normal distribution.

Experiment 2. Accuracy was assessed as the relative proportion of high-value picks in the training phase. A two-tailed one-sample t test showed a significantly higher probability of selecting HV than LV options, $t(22) = 7.24$; $p < .01$; $d_z = 1.51$. In the

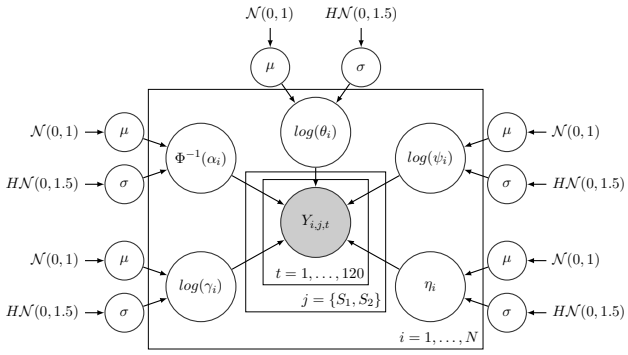


Figure A5. Graphical depiction of the accentuation of differences model. Shaded node represents observed data, unshaded nodes represent model parameters, and distributions without nodes represent priors. $\mathcal{N}$ = Normal distribution; $H\mathcal{N}$ = half-normal distribution.

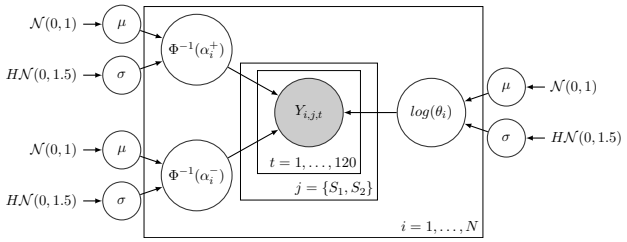


Figure A6. Graphical depiction of the risk-sensitive reinforcement learning model. Shaded node represents observed data, unshaded nodes represent model parameters, and distributions without nodes represent priors. $\mathcal{N}$ = Normal distribution; $H\mathcal{N}$ = half-normal distribution.

experimental phase, the compromise effect was measured as described in the Method section. A two-tailed one-sample t test showed a significant reversal of the compromise effect, $t(22) = 2.49$; $p = .02$; $d_z = 0.52$.

Experiment 3. Accuracy was assessed as the relative proportion of high-value picks in the training phase. A two-tailed one-sample t test showed a significantly higher probability of selecting HV than LV options, $t(22) = 5.82$; $p < .01$; $d_z = 1.21$. In the experimental phase, the attraction effect was measured as described in the Method section. A two-tailed one-sample t test showed a nonsignificant tendency to a reversed attraction effect, $t(22) = 1.26$; $p = .22$; $d_z = 0.26$.

Appendix B

Spektor, Kellen, and Hotaling (2017)

When the Good Looks Bad: An Experimental Exploration of the Repulsion Effect

Mikhail S. Spektor[1], David Kellen[1,2], Jared M. Hotaling[1,3]

[1]University of Basel

[2]University of Syracuse

[3]University of New South Wales

Word count: 952 (Introductions), 1025 (Discussion), 16 (Acknowledgement)

Total words: 1993/2000

References: 35/40

Author Note

Mikhail S. Spektor, Faculty of Psychology, University of Basel, Switzerland. David Kellen, Faculty of Psychology, University of Basel, Switzerland, and College of Arts and Sciences, University of Syracuse, USA. Jared M. Hotaling, Faculty of Psychology, University of Basel, Switzerland, and School of Psychology, University of New South Wales, Australia.

This research was supported by the Center for Economic Psychology and the Center for Cognitive and Decision Sciences, University of Basel, Switzerland, and the Swiss National Science Foundation (Grant 100014_165591 to David Kellen).

Corresponding author: Mikhail S. Spektor, Faculty of Psychology, Missionsstrasse 62a, 4055, Basel, Switzerland. Electronic mail may be sent to michael@spektor.ch

Abstract

When choosing among different options, context seems to play a vital role. For instance, adding a third option can increase the probability of choosing a similar dominating option. This *attraction effect* is one of the most widely studied phenomena in decision-making research. Its prevalence, however, has been recently challenged by the *tainting hypothesis*, according to which the inferior option contaminates the attribute space it is located in, leading to a *repulsion effect*. In an attempt to test the tainting hypothesis and explore the conditions under which dominated options make dominating options look bad, we conducted four (pre-registered) perceptual decision-making studies with a total of 301 participants. We identified two factors influencing individuals' behavior: *stimulus display* and *stimulus design*. Our results contribute to a growing body of literature showing crucial differences between preferential and perceptual decision-making tasks.

Keywords: repulsion effect, attraction effect, context effects, decision making

When humans make decisions, context matters. Several studies have shown that the introduction of a *decoy option* that is similar to but objectively worse than one of the already-available options increases the probability that the similar-but-better option is chosen–an *attraction effect* (e.g., Berkowitsch, Scheibehenne, & Rieskamp, 2014; Gluth, Hotaling, & Rieskamp, 2017; Heath & Chatterjee, 1995; Huber, Payne, & Puto, 1982). For example, in a scenario where one chooses between buying an apple and a banana, the introduction of an equally expensive but less attractive banana will increase the probability that its more attractive counterpart is chosen. In recent years, the attraction effect has played an important role in the comparison of different models of decision making (Bhatia, 2013; Roe, Busemeyer, & Townsend, 2001; Trueblood, Brown, & Heathcote, 2014; Tsetsos, Usher, & Chater, 2010; Usher & McClelland, 2004). This importance derives from the notion that such kind of context effects represent general properties of decision-making behavior. This notion is supported by studies demonstrating contextual effects in perceptual and inferential judgments made by humans and non-human primates (e.g., Parrish, Evans, & Beran, 2015; Trueblood, 2012; Trueblood, Brown, Heathcote, & Busemeyer, 2013). A prominent example is Trueblood et al.'s (2013) demonstration of different context effects–such as the attraction effect–in a perceptual task where individuals were asked to choose the largest of three rectangles (see Choplin & Hummel, 2005, for another perceptual task showing the attraction effects).

Despite the wealth of evidence for the attraction effect, its robustness has recently been challenged in a large-scale replication attempt in which a *repulsion effect* was found to occur just as often (Frederick, Lee, & Baskin, 2014). Repulsion effects are expected under the *tainting hypothesis* (Simonson, 2014, p. 518), according to which similar, yet clearly inferior choice alternatives "taint" the attribute space they are located in (see also Kreps, 1990, p. 28 for a

thought experiment). The repulsion effect has not yet been explored systematically, and the few attempts to observe it have failed to find robust effects (Simonson, 2014, p. 518). Finally, it is currently unclear how the attraction effect is affected by the distance (in the attribute space) between the dominating and dominated options (Soltani, De Martino, & Camerer, 2012; Wedell, 1991).

The present set of four pre-registered studies is an attempt to close these gaps by testing the tainting hypothesis and exploring the conditions under which attraction/repulsion effects are observed. In line with Trueblood et al. (2013), we used a perceptual decision-making task, which provided us with fine-grained control over the features of the stimuli. As reported below, we generally observed large repulsion effects, with an attraction effect only being observed under a very specific set of circumstances. Our investigations showed that two features of our experimental designs, *stimulus design* and *stimulus display,* played a critical role in which context effect we observed. The influence of such features raises concerns regarding the generalizability of context effects to non-preferential choice tasks.

Experiment 1

Previous research has demonstrated that attraction effects disappear when individuals are provided with an unattractive set of options (Huber, Payne, & Puto, 2014, p. 523; Malkoc, Hedgcock, & Hoeffler, 2013). This absence of attraction effects could be due to attribute-space tainting. This possibility highlights the fact that most studies have been conducted along with either positive incentives for the participants (e.g., Herne, 1999) or none whatsoever (e.g., Trueblood et al., 2013). If the occurrence of attraction effects is indeed modulated by the overall attractiveness of the choice context, then one could in principle manipulate it by introducing

negative incentives (i.e., losses), which are well-known to have a disproportionate weight in people's choices (Kahneman & Tversky, 1979).

In Experiment 1 we tested this possible explanation by manipulating monetary incentives. We expected to observe a *gain/loss framing effect* (along the lines of Tversky and Kahneman's, 1981, famous Asian disease problem), comprised of an attraction effect in the context of positive incentives (monetary gains; no tainting of attribute space), and a repulsion effect in the context of negative incentives (monetary losses; tainting of attribute space). This control allowed us to test the tainting hypothesis, but also to explore the moderating role of the attribute distance between the options.

Method

Our main hypotheses, experimental methods, and analysis procedures were pre-registered on the Open Science Framework. Ethical approval was obtained through the institutional review boards of the Faculty of Psychology at the University of Basel (Experiments 1 and 4b) and the College of Arts and Sciences at Syracuse University (Experiments 2, 3, and 4a). The data were partly blinded prior to the analysis. All details on the pre-registration and blinding, raw individual data, and R data-analysis scripts can be found at https://osf.io/4hw6m/.

Participants and procedure. A total of 62 participants (44 female, age 19-55, $M = 25.39$, $SD = 8.37$), mostly students of psychology at the University of Basel, with normal or corrected-to-normal vision participated in Experiment 1. The experiment was conducted in the laboratory with screen resolutions of 1920x1080 pixels. After giving informed consent and filling out the demographic questions, participants completed a calibration task that familiarized them with the response buttons (see Supplemental Materials for details). After completing the calibration task, participants received instructions for the main task and were given three practice trials that were

not part of the main task. On each trial, participants were shown three rectangles of different sizes and had to choose the one with the largest area (see Trueblood et al., 2013). The rectangles were presented in a triangle around the center of the screen, with the vertical positions being jittered across trials. Figure 1 illustrates an example trial. The main task took approximately 45-60 min to complete, including four breaks. Afterwards, participants received the reward accumulated in the main experimental task (CHF 7.10 – 8.80, $M = 8.03$, $SD = 0.36$) in addition to the course-credit equivalent of 1 hour.

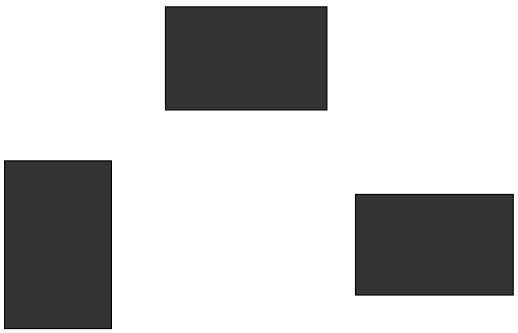


Figure 1. Example of an experimental trial. Participants had to indicate the rectangle with the largest area (in this example the rectangle on the left, narrow/high rectangle). The rectangle in the middle is the wide/low rectangle (WL), and the rectangle on the right is a decoy for WL.

Materials and design. In the main task, participants always saw three rectangles with different area sizes. The two core rectangles differed in orientation, with one being narrow but high (NH; i.e., vertical orientation) and the other wide but low (WL; i.e., horizontal orientation). The third option, the decoy, had the same orientation as one of the core rectangles, but had a smaller area. The option with the unique orientation in each of the trials was the competitor. Note that the repulsion effect is supposed to increase the choice share of the option with the different orientation of the decoy. For notational brevity, we call the option with the decoy's orientation the target, and the other rectangle the competitor, independently of the underlying hypotheses.

Our experimental design consisted of one between-subject manipulation (gain/loss framing) and five within-subject factors. The gain/loss framing concerned whether the task was framed within a context of gains (gain condition: odd participant numbers; 31 participants, 19 female, age 19-48, $M = 25.77$, $SD = 7.74$) or within a context of losses (loss condition: even participant numbers; 31 participants, 25 female, age 19-55, $M = 25.00$, $SD = 9.06$). In the gain condition, participants started with an initial endowment of CHF 0, receiving approximately CHF 0.01 for each correct response (i.e., choosing the rectangle with the largest area), CHF 0.005 for each intermediate response (i.e., choosing the rectangle with the second-largest area), and nothing for each incorrect response (i.e., choosing the rectangle with the smallest area). In the loss condition, participants were endowed with CHF 10 at the beginning and lost nothing, CHF 0.005, and CHF 0.01 for correct, intermediate, and incorrect responses, respectively. At the end of the experiment, participants could receive up to CHF 10 if all responses were correct, and CHF 0 if all responses were incorrect. As such, the incentive structures of both conditions were identical ($M_{\text{gain condition}}$ = CHF 7.97; $M_{\text{loss condition}}$ = CHF 8.09), $t(60) = 1.33$, $p = .187$, $d = 0.34$.

The within-subject factors were set type, difficulty, target option, decoy type, and attribute distance, resulting in a 2x3x2x3x3 within-subject design. Factor "set type" codes which of the core rectangles was larger, WL or NH. Factor "difficulty" codes the difference in areas between the two core rectangles. They differed either by 3%, 7%, or 30% (catch trials) relative to the larger one (i.e., if WL was larger than NH, then the area size of NH was 97%, 93%, and 70% of WL, respectively). Factor "target" codes which of the two core rectangles is the target (i.e., which orientation the decoy has). Factor "decoy type" codes whether the decoy is smaller on the target's weaker attribute (*range decoy*), the target's stronger attribute (*frequency decoy*), or on both attributes (*range-frequency decoy*). This terminology was introduced in the original paper on the attraction effect (Huber et al., 1982). Factor "attribute distance" codes the difference in area between the target and the decoy, which was either 2%, 5%, or 9%. In the catch trials, these differences were 20%, 50%, and 90%, respectively. In total, there were 108 different factor combinations. We created nine unique, symmetrical WL-NH rectangle pairs (230-270 pixels width, 150-190 pixels height) and applied the 108 manipulations to each of them, resulting in a total of 972 trials (for the full trial list, see pre-registration). The main factors of interest are decoy type and attribute distance, while the other factors serve as controls to nullify certain decision strategies (e.g., "pick the unique rectangle" or "take the larger of the two similar ones") and ultimately balance the experimental design. See Figure 2 for an illustration of the within-subject factors.

We used two different instances of dependent variables. When evaluating participants' overall performance in the main task, we considered the proportions of correct, intermediate, and incorrect choices. When testing for the context effects, we relied on the *relative choice share of the target* (RST; Berkowitsch et al., 2014): $\text{RST} = \frac{\Pr(\text{T})}{\Pr(\text{T})+\Pr(\text{C})}$, where Pr(T) is the proportion of

target choices and Pr(C) is the proportion of competitor choices. RST values range from 0 (competitor is always chosen) to 1 (target is always chosen), where RST = .50 indicates an absence of context effects, and RST > .50 and RST < .50 indicates the presence of an attraction and repulsion effect, respectively. By using the RST as a dependent measure, we automatically control for individual prior preferences for horizontally or vertically aligned rectangles. Directional pre-registered hypotheses have been tested with one-tailed tests, where applicable. All other analyses were analyzed using two-tailed tests.

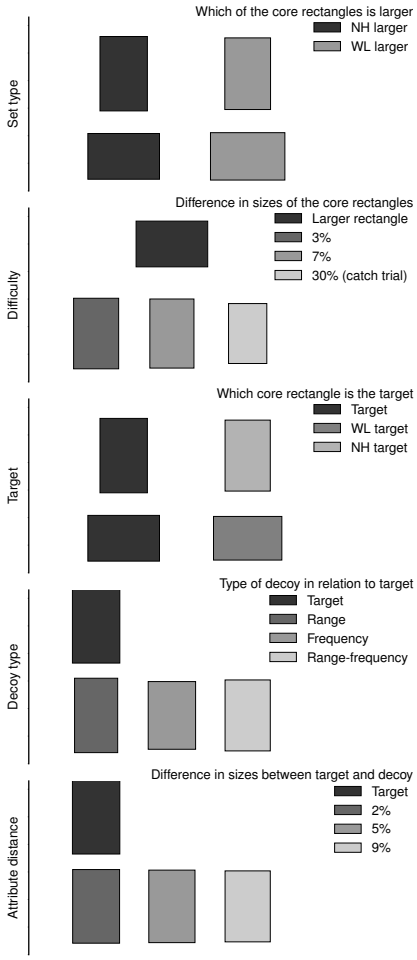


Figure 2. Illustration of the stimuli used across the five within-subject factors in Experiment 1 and 2. WL = wide/low rectangle (i.e., horizontally aligned); NH = narrow/high rectangle (i.e., vertically aligned). See Method/Materials and Design for all explanations of the manipulations.

Results

Data pre-treatment and accuracy checks. Following Trueblood et al. (2013), all of our studies excluded participants according to their overall accuracy in catch trials (in our case, less than 2/3 correct instead of a relative exclusion rule) and trials according to their speed (responses faster than 100ms and longer than 8s). In a second step, we checked whether the order of choice proportions matched the areas of the rectangles. In other words, the largest rectangle should be chosen more often than the second-largest one, which in turn should be chosen more often than the smallest one. Finally, we tested whether trial difficulty influenced choice accuracy such that more difficult trials led to a lower accuracy compared to less difficult trials and catch trials. Overall, we only excluded a small portion of the participants. The remaining participants' responses were generally accurate and tracked the areas of the stimuli. The test results for all four studies are reported in detail in the Supplemental Material.

Confirmatory hypothesis testing. We excluded the catch trials from all hypothesis tests, leaving us with 648 trials per participant. To test for the gain-loss framing effect, we compared the RSTs in the between-subject conditions with a one-tailed t-test for independent samples. **T**he mean RSTs in the gain condition did not differ from their loss-condition counterparts (M = .43, SD = .07, vs. M = .43, SD = .05), $t(60) = -0.14$, $p > .250$, $d = -0.04$. In the loss condition, a large repulsion effect was present as confirmed by a one-tailed, one-sample t-test on RSTs; $t(30) = 7.18$, $p < .001$, $d = 1.29$. In the gain condition, there was no attraction effect, $t(30) = -5.57$, $p > .250$, $d = -1.00$, but another clear repulsion effect (as reflected in the sign of the t value and effect size).

To test for the superiority of the range decoy relative to the other decoy types in the gain condition (replication of Trueblood et al., 2013), we computed a one-tailed repeated-measures t-

test on mean RSTs. The range decoy did not lead to a stronger attraction effect than the other decoy types (M = .43, SD = .07 vs. M = .43, SD = .07), $t < 1$, d = 0.18. Figure 3 (hatched bars) reports the choice proportions for each of the rectangles.

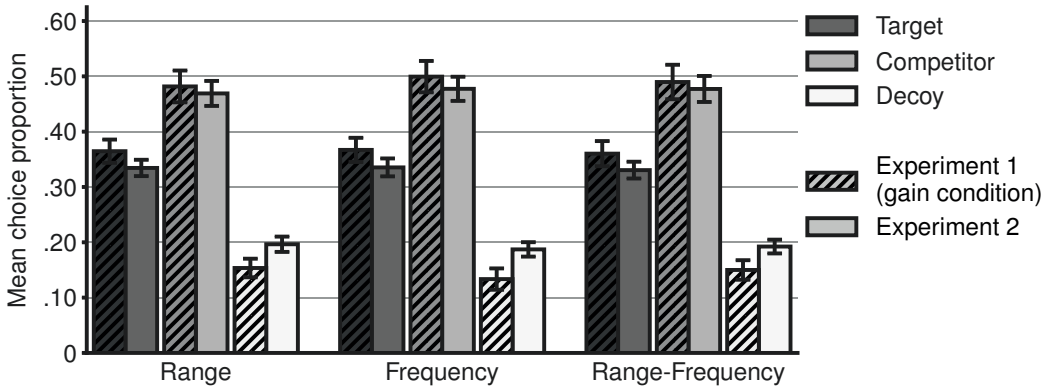


Figure 3. Choice proportions for different decoy types in the gain condition of Experiment 1 (hatched bars) and Experiment 2. For reasons of notational brevity, target always refers to the core rectangle that is similar to the decoy, independently of the underlying hypothesis. Range decoys are weaker on the target's weaker attribute (i.e., are narrower than the target if the target is oriented vertically), frequency decoys are weaker on the target's stronger attribute (i.e., are shorter than the target if the target is oriented vertically), and range-frequency are weaker on both attributes (i.e., are narrower and shorter). Error bars indicate 95% CI.

As our last hypothesis test, we checked for the influence of distance in the attribute space between the target and decoy. We anticipated the possibility of different effects for different conditions, so we performed a 2 (gain/loss framing) x 3 (decoy distance) mixed ANOVA on RSTs. As expected based on the results of the first hypothesis, there was no main effect of gain/loss framing ($F < 1$). There was, however, the predicted main effect of distance, $F(2, 120) = 25.67$, $p < .001$, $\eta_p^2 = .30$. It was characterized by an increase of RSTs (i.e., weakening repulsion

effects), with $M = .41$ ($SD = .08$), $M = .42$ ($SD = .07$), and $M = .45$ ($SD = .06$), for the 2%, 5%, and 9% distances, respectively (see Figure 4, top left panel, for the choice proportions of the individual rectangles). This main effect was independent of gain/loss framing, as corroborated by a non-significant interaction term ($F < 1$).

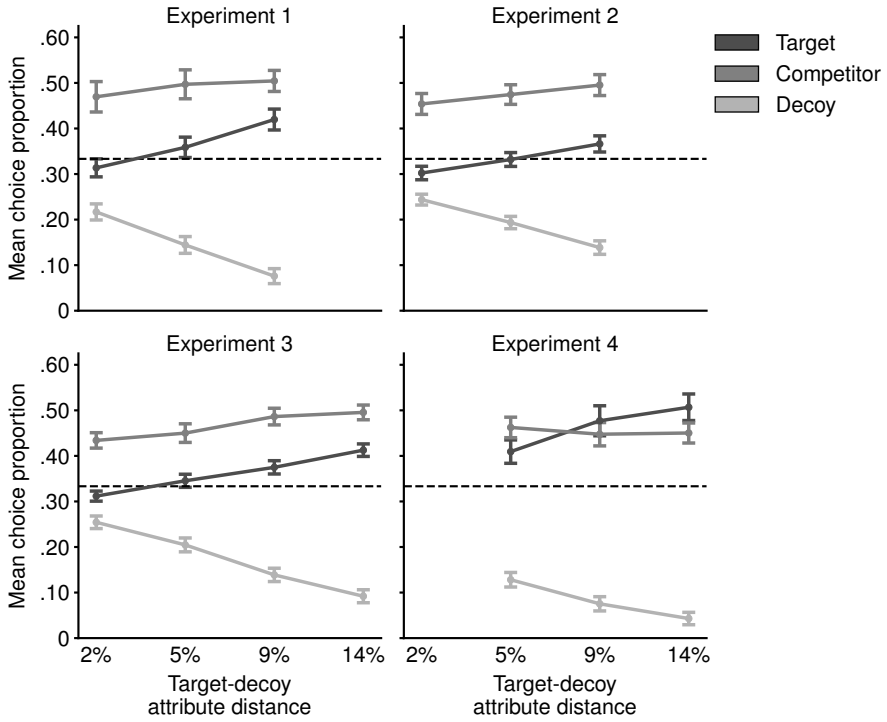


Figure 4. Choice proportions for different target-decoy distances in all experiments. Distances are always in relative area of the target (i.e., 2% indicates that the area of the decoy is 98% of the target's). Experiment 1 includes only the gain-framing condition (as per pre-registration). Experiment 4 includes only the new-trials condition of Experiment 4a, as the direct-replication condition did not manipulate target-decoy distance. Error bars indicate 95% CI.

Exploratory analyses. To assess the robustness of the repulsion effect, we tested the impact of potential covariates. It has been argued that the magnitude of attraction effects increases with deliberation time (Pettibone, 2012). Descriptively, this notion is supported as our participants took longer to respond when they chose the target option (M = 1815ms, Md = 1446ms, SD = 1241ms) than when they chose the decoy (M = 1755ms, Md = 1364ms, SD = 1262ms) or the competitor (M = 1675ms, Md = 1330ms, SD = 1157ms). Another potential covariate is the presence of *strong prior trade-offs*, meaning that individuals have strong preference for one of the core options before the introduction of the decoy (Huber et al., 2014, p. 522). In our paradigm, we have two different levels of prior trade-offs coded into the difficulty: In the easier trials, participants have a stronger "prior preference" for the correct alternative, as it is easier to perceive which of the core rectangles is larger. We checked for both the influence of response time and difficulty and found that both factors influenced the absolute size of the repulsion effect but without a sign flip (i.e., the repulsion effect persists across all factor levels). Since the effects were rather fragile across experiments, we report them together with further analyses in detail in the Supplemental Material.

Experiment 2

The goal of Experiment 2 was to replicate the repulsion effect using a more streamlined design that drops the gain-loss framing manipulation as well as the monetary incentives. The main empirical findings that we aimed to establish were the consistent repulsion effect and its dependency on the attribute distance between the target and decoy.

Method and Results

A total of 61 undergraduate students at Syracuse University with normal or corrected-to-normal vision participated in Experiment 2 (age 18-33, $M = 18.98$, $SD = 2.07$). Besides the gain-loss framing and monetary incentives, the experimental task, procedure, and design was identical to Experiment 1. Participants only received a performance-independent course-credit equivalent of 1 hour.

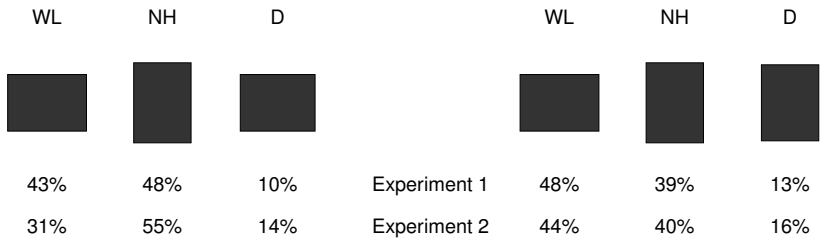


Figure 5. Illustration of two trials sharing the same core rectangles (trial IDs 439 and 443, respectively). Plotted below each rectangle are their corresponding choice proportions for Experiment 1 and 2, respectively. In this case, the decoy rectangle (D) is a range decoy, the wide/low rectangle (WL) is 3% smaller than the narrow/high rectangle (NH), and the distance between the target and D is 5%.

Confirmatory hypothesis testing. We excluded the catch trials from all hypothesis tests, leaving us with 648 trials per participant. A one-tailed, one-sample t-test on RSTs confirmed the same repulsion effect observed in Experiment 1, $t(57) = 8.78$, $p < .001$, $d = 1.15$. The repulsion effect is manifested by means of a relative preference for the competitor over the target. See Figure 5 for an example of such a set. In that example, NH is preferred over WL, when the decoy is similar to WL. On the other hand, if a decoy similar to NH is present, then individuals preferred WL over NH. As the second and last pre-registered hypothesis test, we checked for the

influence of distance in the attribute space between the target and decoy. We performed a repeated-measures ANOVA on RSTs with distance as the within-subject factor. This analysis confirmed the predicted main effect of distance, $F(2, 114) = 5.25$, $p < .01$, $\eta_p^2 = .08$. It was characterized by an increase of RSTs (i.e., weakening repulsion effects), with mean values of .40 ($SD = .08$), .41 ($SD = .08$), and .43 ($SD = .08$), for the 2%, 5%, and 9% distances, respectively (see Figure 4, top right panel, for the choice proportions of the individual rectangles).

Experiment 3

Experiment 3 aimed at making the target-decoy distance more comparable to one used by Trueblood et al. (2013). One notable difference between their design and ours is the distance between targets and decoys in the attribute space. Specifically, they used relative size differences of (on average) 16% and 10% for the range/range-frequency and frequency decoys, respectively. They found that range and range-frequency decoys led to attraction effects, whereas frequency decoys did not produce significant effects (although the observed pattern was in the direction of the attraction effect). In our previous experiments, the largest target-decoy distance was 9%, comparable to Trueblood et al.'s. frequency-decoy attribute distance.

Method and Results

A total of 72 participants, mostly psychology students at Syracuse University, with normal or corrected-to-normal vision participated in Experiment 3 (ages 18-23, $M = 18.98$, $SD = 1.17$). Apart from the following changes, the experimental task, procedure, and design was identical to Experiment 2. The main difference is the addition of a new attribute distance factor level. As a logical progression of our previous factor levels, we added the 14% target-decoy attribute distance level. Consequently, to still maintain the balancing that controls for different

decision strategies, we added a new difficulty factor level for which the core rectangles differed in area sizes by 11%. Instead of the scaled target-decoy distance in the catch trials used in our previous experiments, we fixed it to 20% in Experiment 3. Having seen no differences in target types, we removed this factor and used only range-frequency decoys. The changes resulted in a 2 (set type) x 4 (difficulty) x 2 (target) x 4 (distance) within-subject design. In total, there were 64 different factor combinations within each participant. For each of these combinations we had 9 unique trials, resulting in a total of 576 trials (for the full trial list, see pre-registration). Consequently, the experiment took about 40 minutes to complete and participants received the course-credit equivalent of an hour.

Confirmatory hypothesis testing. We excluded the catch trials from all hypothesis tests, leaving us with 432 trials per participant. A one-tailed, one-sample t-test on RSTs confirmed the repulsion effect in this experiment as well, $t(62) = 9.27$, $p < .001$, $d = 1.17$. As the second and last pre-registered hypothesis test, we checked for the influence of attribute-space distance between the target and the decoy. A repeated-measures ANOVA on RSTs with distance as the within-subject factor confirmed the predicted main effect of distance, $F(3, 186) = 5.76$, $p < .001$, $\eta_p^2 = .08$. It was characterized by an increase of RSTs (i.e., weakening repulsion effects), with mean values of .42 (*SD* = .06), .44 (*SD* = .08), .44 (*SD* = .07), and .45 (*SD* = .06) for the 2%, 5%, 9%, and 14% distances, respectively (see Figure 4, bottom left panel, for the choice proportions of the individual rectangles).

Exploratory analyses. A unique feature of Experiment 3 is that the decoy's area size always differed by 20% from the target's. As a robustness check, we performed an RST analysis using the catch trials (after excluding too inaccurate participants and too fast/slow responses). This analysis is particularly conservative given that decoys were further away from the target in

the attribute space (compared to Trueblood et al., 2013) and that the largest rectangle was more clearly perceivable. Nevertheless, a two-tailed, one-sample t-test on RSTs (M = .49, SD = .03) confirmed the robustness of the repulsion effect; $t(71) = 2.29$, p = .02, d = 0.27.

Experiment 4

Experiment 4 aimed at identifying the moderators that promote the occurrence of attraction/repulsion effects. We identified three factors that might influence whether attraction effects or repulsion effects occur: (i) stimulus design, (ii) stimulus display (i.e., arrangement of the rectangles on-screen), and (iii) absolute size of the rectangles. We believe that the latter's influence is only marginal, resulting in two critical factor combinations that we explored in Experiment 4: a) A stimulus design similar to our previous experiments arranged as in Trueblood et al. (2013), and b) Trueblood et al.'s stimulus design arranged as in our previous experiments.

Method and Results

A total of 83 participants, mostly undergraduate students at Syracuse University, with normal or corrected-to-normal vision participated in Experiment 4a (ages 18-55, M = 19.02, SD = 4.14. Twenty-three psychology students at the University of Basel participated in Experiment 4b (ages 18-30, M = 22.11, SD = 3.38; demographic data of five participants were lost). Apart from the stimulus design, both sub-experiments were identical to Experiments 2 and 3.

In Experiment 4a, we contrasted a *direct-replication condition* (N = 40) with a *new-trials condition* (N = 43) in which we made minimal changes to the way the stimuli were generated. In both conditions, the stimuli were closely arranged along a horizontal line (with some vertical jitter), as done in Trueblood et al. (2013). The new-trials condition contained the very same 180 filler trials as in the direct-replication condition. The remaining 540 experimental trials stemmed

from a 2 (target) x 3 (decoy type) within-subject design in the direct-replication condition (see Trueblood et al., 2013, pp. 903–904 for details) and from a 2 (difficulty: no area difference between the core rectangles as in Trueblood et al., 2013, vs. 7% area difference) x 3 (distance between decoy and target: 5%, 9%, and 14%) x 3 (decoy type) within-subject design in the new-trials condition. Apart from difficulty (one additional level) and decoy distance (two additional levels), the stimulus design of the new-trials condition is identical to the design of the direct-replication condition (i.e., the core rectangles were mostly around 80 pixels x 50 pixels large). With 36 factor combinations, the new-trials condition's design is significantly simpler than the one used in Experiments 1 and 2 (108 factor combinations) and 3 (64 factor combinations), and only slightly more complex than the direct-replication condition (6 factor combinations).

Experiment 4b is a replication of Trueblood et al.'s (2013) attraction-effect experiment, with the sole exception that the options were presented in a triangular arrangement as used in our Experiments 1-3 (see Figure 1). Both experiments took about 40 minutes to complete and participants received the course-credit equivalent of an hour.

Confirmatory hypothesis testing. We excluded the filler trials from all hypothesis tests, leaving us with 540 trials per participant. As expected, we successfully replicated the attraction effect in the direct-replication condition of Experiment 4a (M_{RST} = .55, SD_{RST} = .09), $t(32)$ = 2.92, $p < .01$, d = 0.51. Contrary to our expectations, we did not observe a repulsion effect in the new-trials condition (M_{RST} = .50, SD_{RST} = .07). If anything, participants' behavior tended to go in the direction of the attraction effect, $t(29)$ = 0.33, p = .63, d = 0.06. The difference in mean RSTs between the two conditions was significant; $t(61)$ = 1.98, p = .03, d = 0.50. In contrast to Experiment 4a, we observed a strong repulsion effect in Experiment 4b (M_{RST} = .47, SD_{RST} = .04), as confirmed by a two-tailed one-sample t-test, $t(19)$ = 3.65, $p < .01$, d = 0.82.

Exploratory analyses. We began by checking the influence of distance in the attribute space between the target and decoy in the new-trials condition of Experiment 4a. Specifically, we performed a repeated-measures ANOVA on RSTs with distance as the within-subject factor. As in the previous experiments, this analysis confirmed a main effect of target-decoy attribute distance, $F(2, 58) = 24.71$, $p < .001$, $\eta_p^2 = .46$. This effect was characterized by an increase of RST, with mean values of .47 ($SD = .07$), .51 ($SD = .09$), and .53 ($SD = .07$) for the 5%, 9%, and 14% distances, respectively (see Figure 4, bottom right panel, for the choice proportions of the individual rectangles). In contrast to the previous experiments, this main effect seems to suppress the global RST analysis: Individuals seem to show a repulsion effect for the shortest target-decoy attribute distance, an attraction effect for the largest target-decoy attribute distance, and a null effect for the in-between attribute distance. To confirm this intuition, we ran three two-tailed one-sample t-tests on the RSTs within each distance level separately. Indeed, we observed a small-to-moderate repulsion effect for the 5% distance, $t(29) = 2.31$, $p = .03$, $d = 0.42$, no context effect for the 9% distance, $t < 1$, $d = 0.16$, and a small-to-moderate attraction effect for the 14% distance, $t(29) = 2.10$, $p = .04$, $d = 0.38$. In a next step, we checked for the influence of decoy type separately for each condition of Experiment 4a and Experiment 4b. In none of these cases did decoy type influence the strength of the attraction effect or repulsion effect, all ps $\geq .20$.

A Multiattribute Linear Ballistic Accumulator Account

To gain a more mechanistic understanding of the cognitive process underlying the behavior in our experiments, we fitted the multiattribute linear ballistic accumulator (MLBA) model (Trueblood et al., 2014) to the data of Experiments 3 and 4.

The MLBA describes choice proportions in multiattribute decision tasks by means of five parameters. According to the model, the objective attribute values of the options (in the present case, width and height) are converted into subjective representations. These subjective representations are characterized by a parameter m determining whether individuals prefer options that are very diverse or homogenous with respect to their attributes. Moreover, parameters λ_1 and λ_2 capture the subjective importance that is attributed to positive and negative attribute comparisons, respectively. Individuals' preferential attention toward one of the attributes (e.g., paying more attention to widths than heights) is captured by parameter β. Finally, the model postulates a "normalizing" parameter I_0 that ensures that a decision is reached eventually.

The MLBA was fitted to the individual choices obtained in Experiment 3 and 4 (for details, see the Supplemental Material). As shown in Figure 6, the MLBA was able to provide a close qualitative and quantitative account for the patterns observed in both experiments. In terms of parameter values, we found that the main driving force behind whether repulsion effects or attraction effects are predicted was the ratio of the two λ parameters. When the model predicts attraction effects, the ratios differ only slightly, whereas λ_2 is substantially larger than λ_1 when repulsion effects are predicted.

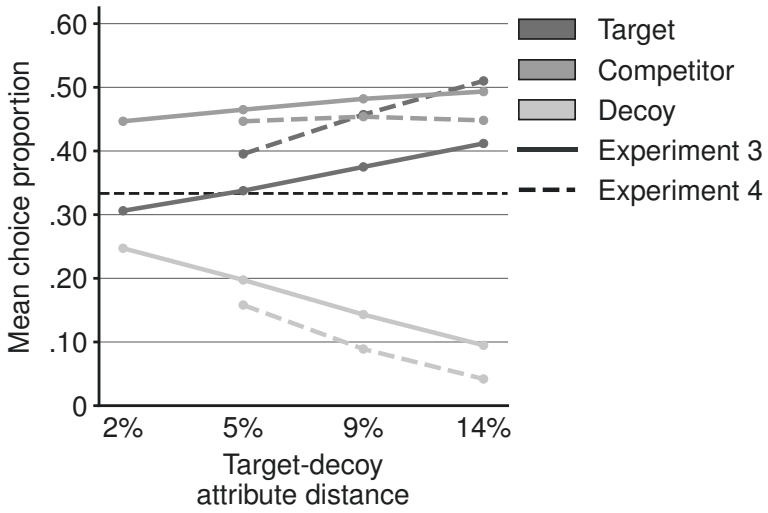


Figure 6. Multiattribute linear ballistic accumulator model predictions for different target-decoy attribute distances in Experiments 3 and 4. Distances are always in relative area of the target (i.e., 2% indicates that the area of the decoy is 98% of the target's). Experiment 4 includes only the new-trials condition of Experiment 4a, as the direct-replication condition did not manipulate target-decoy attribute distance. Lines represent grand-mean predictions.

General Discussion

The present work attempted to test the tainting hypothesis of the repulsion effect according to which unattractive options taint the attribute space they are locating in, thus making other nearby options less attractive. This effect mirrors the attraction effect, in which dominating options appear more attractive. We aimed at determining the conditions under which dominated options yield one of the two effects. Across four pre-registered experiments, we found a large and robust repulsion effect, an effect whose empirical reality had been questioned until recently (e.g., Tsetsos, Chater, & Usher, 2015). Moreover, we also found that both task complexity and

arrangement of options on-screen determine whether attraction effects, null effects, or repulsion effects are observed. Finally, by varying the distance in the attribute space between the dominating option and its dominated counterpart, we found that increases in attribute distance shifted choices towards an attraction effect.

The attraction-repulsion continuum

Since Trueblood et al. (2013), the rectangle-size task has been used in four studies involving humans (Farmer, Warren, El-Deredy, & Howes, 2017; Frederick et al., 2014; Trueblood, Brown, & Heathcote, 2015; Zhen & Yu, 2016). Two of these studies found evidence in favor of the attraction effect, one had mixed results, and one showed a tendency in the direction of the repulsion effect. Our study helps bridging the gap between these results.

We observed two main driving factors: 1) arrangement of the rectangles on-screen, and 2) stimulus design. Surprisingly, the influence of the former far surpasses the latter's: As soon as the options are arranged further apart from each other, the attraction effect disappears entirely and even becomes a robust repulsion effect. But the stimulus design also plays a crucial role: As soon as individuals face choice sets of varying difficulty and, more importantly, with varying attribute distances between the target and decoy, the repulsion effect becomes stronger or the attraction effect becomes weaker, depending on the option arrangement.

Decoys located further away from the target make it easier to notice the dominance relationship between them, whereas closer decoys are at risk to be confused as equally-sized rectangles. In the latter case, individuals might exhibit the *similarity effect* (Tversky, 1972). At first glance, the repulsion effects we demonstrated might seem like a similarity effect, as both predict an increase of choices for the option dissimilar to the decoy. The crucial difference between the similarity effect and the repulsion effect, however, is that the decoy in the former

case is perceived as on par with the target, and as inferior in the latter. In settings like Trueblood et al.'s (2013), where there is no objectively correct choice, it is impossible to tell whether the dominance relationship has been perceived (resulting in a repulsion effect) or whether the target and decoy have been confused with each other (resulting in a similarity effect). However, our experiments controlled for this confound and found no support whatsoever for this "similarity-effect interpretation", given that individuals chose the decoy significantly less often than the target, showing an ability to discriminate between the two (see Supplemental Material).

Interestingly, although the MLBA only predicts repulsion effects under specific circumstances (Tsetsos et al., 2015), it was able to account for the present repulsion effects by placing substantially greater weight on negative comparisons relative to positive comparisons. However, this successful description of the data may stem from the fact that the model was not constrained by having to simultaneously predict other context effects (see Hotaling & Rieskamp, under review, for a demonstration of MLBA's flexibility when fit to only one context effect). A stricter test would require a joint fit of multiple context effects.

The tainting hypothesis

In its present form, the tainting hypothesis simply states that inferior decoys taint the attribute space they are located in. Previous explanations for this tainting included a similarity mechanism (Frederick et al., 2014, p. 493) or the possibility of the target being infected with the decoy's repulsive attributes (Simonson, 2014, p. 518). Both explanations assume higher-level reasoning processes which seem implausible in the rectangle-size task (see Trueblood et al., 2014 for a similar discussion on loss aversion in perceptual tasks). The tainting hypothesis, as we see it, predicts that tainting should be a decreasing function of distance in the attribute space, a

result that was supported empirically. But instead of a framing-dependent tainting, we found universal attribute-space tainting.

The observation of a perceptual context effect that is completely opposite to consumer-choice counterpart is not unique to the present study (for a recent example, see Trueblood & Pettibone, 2017). These gross differences suggest that, despite some obvious structural similarities, there are fundamental differences between the perceptual and preferential tasks usually adopted by researchers (Dutilh & Rieskamp, 2016; Hotaling, Cohen, Shiffrin, & Busemeyer, 2015; Wu, Delgado, & Maloney, 2009). To make matters worse, some of the large behavioral differences observed appear to be due to minor changes in one of the features of the experimental design. Indeed, given the purported generality of the attraction effect across judgment domains, none of these minor changes should have been of any major consequence. Also, neither the MLBA nor any of the extant models provides any a priori mechanisms for *explaining* these behavioral differences. Altogether, it seems unwise to assume by default that choices in a perceptual task are proxies for preferential decision making. We conclude that in the rectangle-size task, researchers are much more likely to observe a repulsion than an attraction effect, as the latter requires the joint occurrence of several specific factor combinations, and the former arises in all other cases.

Conclusion

The observation of a robust repulsion effect has implications for current theoretical discussions. The overall performance of different theories has been assessed in terms of their ability to simultaneously account for different context effects observed in the literature, but also in terms of their ability to predict unobserved effects (Tsetsos, Chater, & Usher, 2015; see also Roberts & Pashler, 2000). Until now, the prediction of a repulsion effect has been perceived as

an unfavorable feature for a model to have due to the lack of empirical support. The present demonstration that the repulsion effect is a real, robust, and replicable phenomenon changes that.

Research disclosure statements:

The total number of excluded observations and the reasons for making these exclusions have been reported in the Method section. All independent variables and manipulations, whether successful or failed, have been reported in the Method section (all experiments) and pre-registered before collecting data (Experiments 1-4a). All dependent variables or measures that were analyzed for this article's target research question have been reported in the Method section (all experiments) and pre-registered before collecting data (Experiments 1-4a).

Sample size:

In Experiment 1, given that attraction effects are typically very robust and strong, achieving even a weak reversal of the attraction effect in the loss condition would have led to a large effect size in a between-subject comparison. Given our sample size, $1\text{-}\beta = .80$, $\alpha = .05$, and a one-tailed test, effect sizes of $d = 0.65$ were detectable. The rest of the analyses rely on within-subject effects, for which we have almost 1,000 observations per participant, plenty for within-subject analyses. The main effect of interest, the repulsion effect, had an effect size of $d > 1.00$. For a power of $1\text{-}\beta = .95$, only 13 participants would have been required. With our sample sizes, we were able to detect effects as small as $d = 0.43$.

As we pre-registered our experiments, we did not have any optional stopping rules (except for Experiment 4a, for which we pre-registered the optional stopping rule). Participants exceeding the pre-registered sample sizes were allowed to participate because they registered for participation before the last participant required to fulfill the pre-registration participated (2 in Experiment 1, 1 in Experiment 2, 12 in Experiment 3).

Author contributions:

All authors developed the study concept and design. D. Kellen and M. S. Spektor coordinated the data collection. M. S. Spektor performed the data analysis and interpretation under supervision of D. Kellen and J. M. Hotaling. J. M. Hotaling performed computational modeling on Experiments 3 and 4. M. S. Spektor drafted the manuscript, D. Kellen critically revised it, and J. M. Hotaling gave final critical comments. All authors approved the final version of the manuscript for submission.

Acknowledgements:

We thank Markus Steiner, Tris Buck, Tehilla Mechera-Ostrovsky, Henrik Singmann, and Sebastian Gluth for their help.

References

Berkowitsch, N. A. J., Scheibehenne, B., & Rieskamp, J. (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General*, *143*, 1331–1348. https://doi.org/10.1037/a0035159

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, *120*, 522–543. https://doi.org/10.1037/a0032457

Choplin, J. M., & Hummel, J. E. (2005). Comparison-induced decoy effects. *Memory & Cognition*, *33*, 332–343. https://doi.org/10.3758/BF03195321

Dutilh, G., & Rieskamp, J. (2016). Comparing perceptual and preferential decision making. *Psychonomic Bulletin & Review*, *23*, 723–737. https://doi.org/10.3758/s13423-015-0941-1

Farmer, G. D., Warren, P. A., El-Deredy, W., & Howes, A. (2017). The effect of expected value on attraction effect preference reversals. *Journal of Behavioral Decision Making*, *30*, 785–793. https://doi.org/10.1002/bdm.2001

Frederick, S., Lee, L., & Baskin, E. (2014). The limits of attraction. *Journal of Marketing Research*, *51*, 487–507. https://doi.org/10.1509/jmr.12.0061

Gluth, S., Hotaling, J. M., & Rieskamp, J. (2017). The attraction effect modulates reward prediction errors and intertemporal choices. *Journal of Neuroscience*, *37*, 371–382. https://doi.org/10.1523/JNEUROSCI.2532-16.2017

Heath, T. B., & Chatterjee, S. (1995). Asymmetric decoy effects on lower-quality versus higher-quality brands: Meta-analytic and experimental evidence. *Journal of Consumer Research*, *22*, 268–284. https://doi.org/10.1086/209449

Herne, K. (1999). The effects of decoy gambles on individual choice. *Experimental Economics*, *2*, 31–40. https://doi.org/10.1023/A:1009925731240

Hotaling, J. M., Cohen, A. L., Shiffrin, R. M., & Busemeyer, J. R. (2015). The dilution effect and information integration in perceptual decision making. *PLOS ONE*, *10*, 1–19. https://doi.org/10.1371/journal.pone.0138481

Hotaling, J. M., & Rieskamp, J. (under review). A quantitative test of sequential sampling models of multialternative context effects.

Huber, J., Payne, J. W., & Puto, C. P. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*, 90–98. https://doi.org/10.1086/208899

Huber, J., Payne, J. W., & Puto, C. P. (2014). Let's be honest about the attraction effect. *Journal of Marketing Research*, *51*, 520–525. https://doi.org/10.1509/jmr.14.0208

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–292. https://doi.org/10.2307/1914185

Kreps, D. M. (1990). *A Course in Microeconomic Theory*. Princeton, NJ: Princeton University Press.

Malkoc, S. A., Hedgcock, W., & Hoeffler, S. (2013). Between a rock and a hard place: The failure of the attraction effect among unattractive alternatives. *Journal of Consumer Psychology*, *23*, 317–329. https://doi.org/10.1016/j.jcps.2012.10.008

Parrish, A. E., Evans, T. A., & Beran, M. J. (2015). Rhesus macaques (Macaca mulatta) exhibit the decoy effect in a perceptual discrimination task. *Attention, Perception, & Psychophysics*, *77*, 1715–1725. https://doi.org/10.3758/s13414-015-0885-6

Pettibone, J. C. (2012). Testing the effect of time pressure on asymmetric dominance and compromise decoys in choice. *Judgment and Decision Making*, *7*, 513–521.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing.

Psychological Review, *107*, 358–367. https://doi.org/10.1037/0033-295X.107.2.358

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review*, *108*, 370–392. https://doi.org/10.1037/0033-295X.108.2.370

Simonson, I. (2014). Vices and virtues of misguided replications: The case of asymmetric dominance. *Journal of Marketing Research*, *51*, 514–519. https://doi.org/10.1509/jmr.14.0093

Soltani, A., De Martino, B., & Camerer, C. (2012). A range-normalization model of context-dependent choice: A new model and evidence. *PLoS Computational Biology*, *8*, 1–15. https://doi.org/10.1371/journal.pcbi.1002607

Trueblood, J. S. (2012). Multialternative context effects obtained using an inference task. *Psychonomic Bulletin & Review*, *19*, 962–968. https://doi.org/10.3758/s13423-012-0288-9

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, *121*, 179–205. https://doi.org/10.1037/a0036137

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2015). The fragile nature of contextual preference reversals: Reply to Tsetsos, Chater, and Usher (2015). *Psychological Review*, *122*, 848–853. https://doi.org/10.1037/a0039656

Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers: Context effects are fundamental to decision making. *Psychological Science*, *24*, 901–908. https://doi.org/10.1177/0956797612464241

Trueblood, J. S., & Pettibone, J. C. (2017). The phantom decoy effect in perceptual decision making. *Journal of Behavioral Decision Making*, *30*, 157–167.

https://doi.org/10.1002/bdm.1930

Tsetsos, K., Chater, N., & Usher, M. (2015). Examining the mechanisms underlying contextual preference reversal: Comment on Trueblood, Brown, and Heathcote (2014). *Psychological Review*, *122*, 838–847. https://doi.org/10.1037/a0038953

Tsetsos, K., Usher, M., & Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological Review*, *117*, 1275–1293. https://doi.org/10.1037/a0020580

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299. https://doi.org/10.1037/h0032955

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458. https://doi.org/10.1126/science.7455683

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*, 757–769. https://doi.org/10.1037/0033-295X.111.3.757

Wedell, D. H. (1991). Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 767–778. https://doi.org/10.1037//0278-7393.17.4.767

Wu, S.-W., Delgado, M. R., & Maloney, L. T. (2009). Economic decision-making compared with an equivalent motor task. *Proceedings of the National Academy of Sciences*, *106*, 6088–6093. https://doi.org/10.1073/pnas.0900102106

Zhen, S., & Yu, R. (2016). The development of the asymmetrically dominated decoy effect in young children. *Scientific Reports*, *6*, 1–7. https://doi.org/10.1038/srep22678

Supplemental material for:

When the Good Looks Bad: An Experimental Exploration of the Repulsion Effect

Mikhail S. Spektor[1], David Kellen[1,2], Jared M. Hotaling[1,3]

[1]University of Basel

[2]University of Syracuse

[3]University of New South Wales

Experiment 1

Calibration task

In the calibration task, participants saw a square on either the right, top, or left side of the screen and had to indicate the position of rectangle using the arrow keys (left, up, right, respectively) as quickly as possible without sacrificing accuracy. After twenty warm-up trials, participants completed ten calibration trials. If they answered correctly on fewer than nine calibration trials, another 10-trial block was added.

Applying exclusion criteria

No participants were excluded based on their accuracy on catch trials (accuracy = proportion of correct responses, $M = .98$, $SD = .03$). We excluded one too-fast response (< 0.01%) and 516 too-slow responses (1.28%). Qualitatively, results do not change when analyzed with all data. In a second step, we checked whether the order of choice proportions matched the areas of the rectangles. In other words, the largest rectangle should be chosen more often than the second-largest one, which in turn should be chosen less often than the smallest one. A repeated-measures ANOVA on the mean choice proportions of each participant confirmed that participants' choices matched the areas of the rectangles; $F(2, 180) = 1440.02$, $p < .001$, $\eta_p^2 = .94$. As a final check, we tested whether trial difficulty influenced choice accuracy such that more difficult trials led to a lower accuracy compared to more difficult trials and catch trials. A repeated-measures ANOVA on the mean choice proportions of each participant across the different difficulty levels confirmed this, $F(2, 180) = 247.76$, $p < .001$, $\eta_p^2 = .73$. See Figure S1 (left panel) for choice proportions for each difficulty level aggregated across experiments.

Further analyses

To check for both the influence of response time (RT) and difficulty, we performed a median split on the RTs within each participant and ran a 2 (gain/loss framing) x 3 (decoy distance) x 2 (difficulty) x 2 (RT) between-within-subject ANOVA on RSTs. The effect of decoy distance persisted despite the additional covariates, $F(2, 120) = 27.06$, $p < .001$, $\eta_p^2 = .31$. Of the other effects, only the main effect of RTs, $F(1, 60) = 52.84$, $p < .001$, $\eta_p^2 = .46$, and the interaction of RTs and difficulty, $F(1, 60) = 14.36$, $p < .001$, $\eta_p^2 = .19$, were significant. The main effect of RTs was characterized by higher RSTs when responses were slow ($M = .46$, $SD = .10$) compared to fast responses ($M = .40$, $SD = .10$). The interaction pattern manifested as a positive main effect of RT, but a differing effect of difficulty depending on RT: For fast responses, difficult trials had higher mean RSTs ($M = .41$, $SD = .10$) than easier trials ($M = .38$, $SD = .10$). For slow responses, difficult trials had lower mean RSTs ($M = .45$, $SD = .10$) than easier trials ($M = .47$, $SD = .09$). In all cases, RSTs remained below .50, supporting the notion of a consistent and robust repulsion effect, not a within-subject mixture of attraction effects in some conditions (e.g., fast responses in difficult trials) and (stronger) repulsion effects in others.

We also found little inter-individual differences in the choice rates, with only 6% of the participants choosing the target more often than the competitor, while the remaining 94% preferred the competitor. This consistency across individuals should alleviate any concerns that the reported repulsion effects are the byproduct of the aggregating heterogeneous individual data (see Liew, Howe, & Little, 2016; Simonson, 2014).

We assessed the sensitivity of individuals to changes in the stimulus design in two different ways. In a first analysis, we investigated whether increasing attribute distances between targets and decoys lead to higher choices of targets relative to decoys (i.e., analysis

corresponding to Figure 4). This intuition was confirmed by a one-way repeated-measures ANOVA on the difference of choice proportions between target and decoy, $F(2, 122) = 392.81$, $p < .001$, $\eta_p^2 = .87$. In a second analysis, we checked whether increasing attribute distances also increased led to more accurate choices in general (see Figure S1, right panel, for choice proportions for each target-decoy attribute distance level aggregated across experiments). This was also confirmed by a one-way repeated-measures ANOVA on choice accuracy, $F(2, 122) = 134.61$, $p < .001$, $\eta_p^2 = .69$.

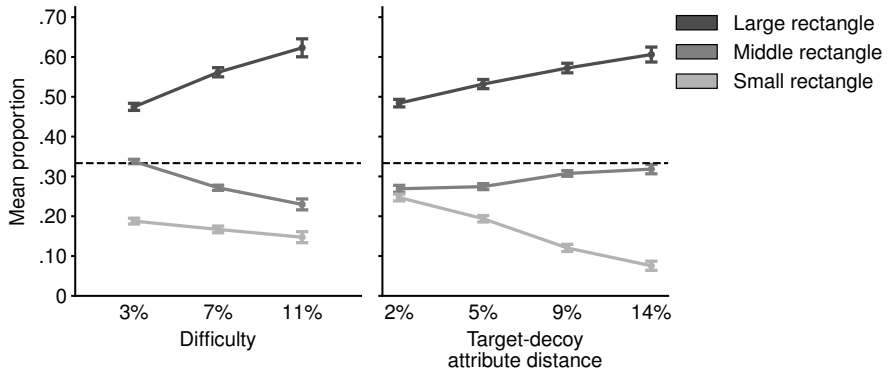


Figure S1. Choice proportions for different difficulties and target-decoy attribute distances aggregated across all experiments with these factors. Distances are always in relative area of the larger rectangle (i.e., 2% target-decoy attribute distance indicates that the area of the decoy is 98% of the target's). Error bars indicate 95% CI.

Experiment 2

Applying exclusion criteria

We used the same exclusion criteria as in Experiment 1. Three participants (5%) were excluded based on their accuracy on catch trials (accuracy = proportion of correct responses, *M* = .89, *SD* = .10). We excluded 133 too-fast responses (0.24%) and 92 too-slow responses (0.16%). Qualitatively, results do not change when analyzed with all data. Following the analysis procedure from Experiment 1, we checked whether the order of choice proportions matched the areas of the rectangles. A repeated-measures ANOVA on the mean choice proportions of each participant confirmed this intuition, $F(2, 168) = 595.81$, $p < .001$, $\eta_p^2 = .88$. We tested whether trial difficulty influenced choice accuracy such that more difficult trials led to a lower accuracy compared to more difficult trials and catch trials. A repeated-measures ANOVA on the mean choice proportions of each participant across the different difficulty levels confirmed this, $F(2, 168) = 157.68$, $p < .001$, $\eta_p^2 = .65$.

Further analyses

For Experiment 2, we repeated the same exploratory analyses as in Experiment 1. A 3 (decoy distance) x 2 (difficulty) x 2 (RT) within-subject ANOVA on RSTs did not confirm the incidental findings observed in the previous experiment. While the effect of distance persisted, $F(2, 114) = 5.41$, $p < .01$, $\eta_p^2 = .09$, both the main effect of RT, $F(1, 57) = 1.62$, $p = .21$, $\eta_p^2 = .03$, as well as the interaction between RT and difficulty disappeared ($F < 1$). Only one participant chose the target more often than the competitor (compared to 57 in the opposite direction), again easing concerns about potential aggregation effects.

We assessed the sensitivity of individuals to changes in the stimulus design in the same two ways as in Experiment 1. In a first analysis, we investigated whether increasing attribute

distances between targets and decoys lead to higher choices of targets relative to decoys. This was confirmed by a one-way repeated-measures ANOVA on the difference of choice proportions between target and decoy, $F(2, 114) = 107.49$, $p < .001$, $\eta_p^2 = .65$. In a second analysis, we checked whether increasing attribute distances also led to more accurate choices in general. This was also confirmed by a one-way repeated-measures ANOVA on choice accuracy, $F(2, 114) = 63.82$, $p < .001$, $\eta_p^2 = .53$.

Experiment 3

Applying exclusion criteria

Experiment 3 used the same exclusion criteria as in Experiments 1 and 2. Nine participants (12.5%) were excluded based on their accuracy on catch trials (accuracy = proportion of correct responses, $M = .85$, $SD = .15$). We excluded 220 too-fast responses (0.61%) and 44 too-slow responses (0.12%). Qualitatively, results do not change when analyzed with all data. Following the analysis procedure from our previous experiments, we checked whether the order of choice proportions matched the areas of the rectangles. A repeated-measures ANOVA on the mean choice proportions of each participant confirmed this intuition, $F(2, 183) = 335.34$, $p < .001$, $\eta_p^2 = .79$. We tested whether trial difficulty influenced choice accuracy such that more difficult trials led to a lower accuracy compared to more difficult trials and catch trials. A repeated-measures ANOVA on the mean choice proportions of each participant across the different difficulty levels confirmed this, $F(3, 244) = 71.51$, $p < .001$, $\eta_p^2 = .47$.

Further analyses

A 4 (decoy distance) x 3 (difficulty) x 2 (RT) within-subject ANOVA on RSTs did not consistently confirm the incidental findings observed in Experiment 1. While the effect of

distance persisted, $F(3, 186) = 5.12$, $p < .01$, $\eta_p^2 = .08$, only the main effect of RT, $F(1, 62) = 4.21$, $p = .04$, $\eta_p^2 = .06$ persisted. The interaction between RT and difficulty disappeared; $F(2, 124) = 2.58$, $p = .08$, $\eta_p^2 = .04$. Instead, an interaction between RT and distance emerged; $F(3, 186) = 3.33$, $p = .02$, $\eta_p^2 = .05$. This interaction was characterized by a stronger increase of RSTs with increasing distance for slower RTs compared to faster RTs. For fast RTs, mean RSTs were .42 (*SD* = .15), .42 (*SD* = .15), .43 (*SD* = .12), and .43 (*SD* = .14) for the 2%, 5%, 9%, and 14% target-decoy distances, respectively. For slow responses, they were .42 (*SD* = .14), .45 (*SD* = .14), .44 (*SD* = .13), and .48 (*SD* = .14). Only five participants (8%) chose the target more often than the competitor (compared to 58 participants in the opposite direction), again easing concerns about potential aggregation effects.

We assessed the sensitivity of individuals to changes in the stimulus design in the same two ways as in Experiments 1 and 2. In a first analysis, we investigated whether increasing attribute distances between targets and decoys lead to higher choices of targets relative to decoys. This was confirmed by a one-way repeated-measures ANOVA on the difference of choice proportions between target and decoy, $F(3, 186) = 169.24$, $p < .001$, $\eta_p^2 = .73$. In a second analysis, we checked whether increasing attribute distances also increased led to more accurate choices in general. This was also confirmed by a one-way repeated-measures ANOVA on choice accuracy, $F(3, 186) = 51.12$, $p < .001$, $\eta_p^2 = .45$.

Experiment 4

Experiment 4b is the only experiment that has not been pre-registered prior to data collection. This experiment completes the hypothetical design matrix of differences between our and Trueblood et al.'s (2013) experiments mentioned in Section Experiment 4 of the main manuscript. It has not been pre-registered as its sole purpose is to serve as a proof of concept. Data for this experiment were collected in parallel to data collection of Experiment 4a and, similarly to the pre-registered experiments, are made available on the same OSF project.

Applying exclusion criteria

We used the same exclusion criteria as in the previous experiments. Note that the difficulty of the filler trials in Experiment 4 was higher (average area difference between largest rectangle and second-largest rectangle is 20%) than in the catch trials used in Experiments 1 to 3 (28% average area difference between largest rectangle and second-largest rectangle). Twenty participants (24.1%) were excluded based on their accuracy on filler trials in Experiment 4a (accuracy = proportion of correct responses, $M = .79$, $SD = .20$). We excluded 321 too-fast responses (0.71%) and 320 too-slow responses (0.71%). The results of one hypothesis change qualitatively when analyzed with all data and will be reported in the respective subsection. All other results are qualitatively unaffected by the exclusion. Due to a computer crash, the data of one participant of Experiment 4b were lost. Of the remaining 22 participants, two participants (9.1%) were excluded due to low accuracy on filler trials ($M = .81$, $SD = .12$). We excluded 8 too-fast responses (0.06%) and 76 too-slow responses (0.53%). Results do not change qualitatively when analyzed with all data.

Further analyses

As a follow-up analysis to the one reported in the main manuscript, we ran a 3 (decoy distance) x 2 (difficulty) x 2 (RT) within-subject ANOVA on RSTs. In this analysis, the main effect of distance became slightly stronger, $F(2, 58) = 28.43$, $p < .001$, $\eta_p^2 = .49$, and the only other significant effect was the main effect of RT, $F(1, 29) = 18.71$, $p < .001$, $\eta_p^2 = .39$; all other *p*s > .45. The main effect of RT is characterized by a repulsion effect for fast responses ($M = .46$, $SD = .13$) and an attraction effect for slow responses ($M = .55$, $SD = .11$). In a next step, we checked for the influence of decoy type separately for each condition of Experiment 4a and Experiment 4b. In none of these cases did decoy type influence the strength of the attraction effect or repulsion effect, all *p*s ≥ .20.

We assessed the sensitivity of individuals to changes in the stimulus design in the new-trials condition of Experiment 4a in the same two ways as in the previous experiments. First, we investigated whether increasing attribute distances between targets and decoys lead to higher choices of targets relative to decoys. This was confirmed by a one-way repeated-measures ANOVA on the difference of choice proportions between target and decoy, $F(2, 58) = 152.23$, $p < .001$, $\eta_p^2 = .84$. In a second analysis, we checked whether increasing attribute distances also increased led to more accurate choices in general. This was also confirmed by a one-way repeated-measures ANOVA on choice accuracy, $F(2, 58) = 10.34$, $p < .001$, $\eta_p^2 = .26$.

Analyses across experiments

To rule out that the repulsion effects we observed resulted (solely) from a confusion of the target with the decoy (i.e., a similarity effect; Tversky, 1972), we assessed whether individuals were able to discriminate the two. To do so, we looked at the most-difficult-to-discriminate factor level of target-decoy attribute distance (2%) across all experiments that included that factor level (Experiments 1 to 3). A two-tailed repeated-measures t-test on the difference of the choice proportions of targets (M = .31, SD = .05) and decoys (M = .24, SD = .05) confirmed that our participants were able to make that distinction, $t(182) = 14.96$, $p < .001$, $d = 1.11$.

As a last analysis, we checked participants' bias towards horizontally or vertically aligned rectangles. Across all our experiments, individuals had a preference for horizontally aligned rectangles, as confirmed by a two-tailed repeated-measures t-test, $t(265) = 9.64$, $p < .001$, $d = 0.59$ (see Table S1 for choice proportions for each type of rectangle in the individual experiments). Descriptively, participants in Trueblood and colleagues' (2013) attraction effect experiment showed the same preference, albeit not reaching significance, $t(48) = 1.05$, $p = .30$, $d = 0.15$.

Table S1

Choice proportions for each type of rectangle in all our experiments and in TBHB2013

Experiment	*N*	Wide/low rectangle (horizontal orientation)	Narrow/high rectangle (vertical orientation)	Decoy (both orientations)
TBHB2013	49	.49 (.17)	.44 (.17)	.06 (.04)
1	62	.52 (.09)	.38 (.08)	.10 (.03)
2	58	.46 (.08)	.41 (.07)	.14 (.04)
3	63	.46 (.07)	.40 (.07)	.14 (.04)
4a (new trials)	30	.62 (.12)	.30 (.11)	.07 (.04)
4a (replication)	33	.49 (.11)	.45 (.11)	.06 (.04)
4b	20	.55 (.11)	.35 (.10)	.10 (.04)

Notes. TBHB2013 = Attraction effect experiment reported in Trueblood, Brown, Heathcote, & Busemeyer (2013). *N* = sample size after applying the respective exclusion criteria. New trials = new-trials condition of Experiment 4a. Replication = direct-replication condition of Experiment 4a. Cells depict means with standard deviations in parentheses. Some rows do not add up to 1 due to rounding.

Modeling details

To fit the model, we used the code provided by Trueblood, Brown, and Heathcote (2014) on log-transformed widths and heights. We modified the code slightly so as to avoid the crashes that would otherwise occur when the maximum input to the accumulators was not above 0. We used the Nelder-Mead method as implemented in Matlab's `fminsearch` to find the maximum likelihood estimates for each individual separately. We restarted the fitting procedure 108 times with random starting values to make sure that we found the global maximum. Descriptive statistics of the parameter estimates are reported in Table S2.

Table S2

Summary statistics for the parameter estimates of the MLBA model in Experiments 3 and 4

Experiment	β	m	I_0	λ_1	λ_2
3	1.90 (3.13)	1.36 (5.98)	8.83 (7.14)	0.07 (0.15)	0.41 (0.35)
4a (new trials)	0.80 (0.55)	2.06 (2.31)	6.35 (7.10)	0.06 (0.04)	0.12 (0.09)
4a (replication)	1.46 (2.29)	1.43 (1.93)	5.61 (6.44)	0.09 (0.05)	0.13 (0.09)
4b	0.65 (1.28)	2.36 (2.53)	12.67 (7.20)	0.03 (0.03)	0.10 (0.08)

Notes. MLBA = Multiattribute linear ballistic accumulator. New trials = new-trials condition of Experiment 4a. Replication = direct-replication condition of Experiment 4a. Cells depict means with standard deviations in parentheses.

References

Liew, S. X., Howe, P. D. L., & Little, D. R. (2016). The appropriacy of averaging in the study of context effects. *Psychonomic Bulletin & Review*, *23*, 1639–1646. https://doi.org/10.3758/s13423-016-1032-7

Simonson, I. (2014). Vices and virtues of misguided replications: The case of asymmetric dominance. *Journal of Marketing Research*, *51*, 514–519. https://doi.org/10.1509/jmr.14.0093

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, *121*, 179–205. https://doi.org/10.1037/a0036137

Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers: Context effects are fundamental to decision making. *Psychological Science*, *24*, 901–908. https://doi.org/10.1177/0956797612464241

Tversky, A. (1972). Choice by elimination. *Journal of Mathematical Psychology*, *9*, 341–367. https://doi.org/10.1016/0022-2496(72)90011-9

Appendix C

Gluth, Spektor, and Rieskamp (2017)

Value-based attentional capture affects multi-alternative decision making

Sebastian Gluth*+, Mikhail S. Spektor* & Jörg Rieskamp

Department of Psychology, University of Basel, Basel, Switzerland

*Equal contribution

+Corresponding author: Sebastian Gluth
Department of Psychology, University of Basel
Missionsstrasse 62a, 4055 Basel, Switzerland
Phone: +41 61 2070606
Email: sebastian.gluth@unibas.ch

Number of words in main text: 1240
Number of references: 11
Number of display items: 3

Recently, studying choices between more than two alternatives has attracted growing interest in decision neuroscience. In such choice settings, the relative preference between two options is often influenced by a third option, which violates the economic principle of independence and has fundamental implications for the neurocognitive principles of decision making[1–4]. However, two recent studies disagreed on whether the third option's value decreases or increases choice accuracy[1,2]. We sought to clarify this contradiction by replicating and extending the study of Chau and colleagues[2] (henceforth Chau2014). To our surprise, we did not observe the positive influence of better third options on choice accuracy reported by Chau2014 in any of three behavioral and one eye-tracking experiments with a total of 147 participants. Instead, we propose an alternative account of the third option's impact on decision making in the Chau2014 paradigm: value-based attentional capture[5].

In Experiment 1, we tested $n_1 = 31$ participants using the Chau2014 paradigm (**Fig. 1a,b**). Chau2014 reported that *relative choice accuracy*, that is, the probability of choosing the high-value (HV) over the low-value (LV) option, increased if the value of a third, unavailable distractor option (D) was comparatively high (i.e., when the value difference HV-D was low). Initially, we hypothesized that the presence of two attributes (i.e., reward probability and magnitude) may drive this effect (for details, see **Supplementary Methods**). Therefore, we added a set of "novel trials" (**Fig. 1c**) that allowed disentangling the predictions of the Chau2014 model, our initial hypothesis, so-called "context effects" such as the attraction[4,6] and phantom decoy effects[7], and value-based attentional capture (**Fig. 1d**). First, we analyzed decisions made in the (non-novel) trials that were identical to those used by Chau2014. Although the overall performance in these trials was not different from the performance in Chau2014 (**Supplementary Fig. 1** and **Supplementary Table 1**), we did not observe the negative effect of HV-D on relative choice accuracy reported in Chau2014 (**Table 1**). Remarkably, the analysis of the novel trials revealed a main effect of Dominance that ran counter to the effect of Chau2014 (**Supplementary Fig. 2a**).

Given these surprising results, we conducted further analyses to understand whether and how D influences the decision process. First, we analyzed *absolute choice accuracy*, that is, the probability of choosing HV when including all trials. We obtained a positive, albeit non-significant effect of HV-D (**Table 1**). Second, we found that the main effect of Dominance in the novel trials was even more pronounced for absolute choice accuracy (**Supplementary Fig. 2b**). Third, we regressed the propensity for choosing D itself on D's value, which revealed a strong positive effect (**Table 1**). Fourth, we analyzed response times (RTs) of HV and LV choices and found that higher values of D slowed down RTs (**Table 1**). Importantly, these effects are neither predicted by the Chau2014 model nor by the divisive normalization model[1], which focuses on relative choice accuracy and does not make RT predictions. However, the concept of value-based attentional capture[5], which states that (even irrelevant) value-laden stimuli attract attention and impair goal-directed actions, can explain the results. To substantiate this explanation, we conducted two additional experiments: In Experiment 2, we investigated whether increasing attentional capacity (via increasing the deliberation time) reduces the influence of D's value. In Experiment 3, we recorded participants' eye movements to assess whether high-value Ds receive more attention and thereby impair choice accuracy.

Experiment 2 comprised two groups (n_{2A}/n_{2B} = 25/24). Group A participants conducted the task as in Experiment 1, but Group B participants had 6 rather than 1.5 s to make decisions. For Group A, there was again no negative HV-D effect on relative choice accuracy (neither for Group B), but there was a significantly positive HV-D effect on absolute choice accuracy (**Table 1**). As predicted, this effect was absent and significantly lower in Group B, as was the influence of the mere presence of D (**Fig. 2a**). For the novel trials, Group B exhibited a main effect of Similarity (**Fig. 2b**), consistent with combined attraction and phantom-decoy effects (compare with predictions shown in **Fig. 1d**) and with previous research showing that such effects require longer deliberation times[8]. In summary,

Experiment 2 revealed that the negative influence of the value of D on absolute choice accuracy depends on the exertion of time pressure in the Chau2014 task paradigm.

In Experiment 3 ($n_3 = 23$), we used eye tracking in combination with the paradigm of Experiments 1 and 2A. As before, the negative HV-D effect reported in Chau2014 was not significant (**Table 1**). A path analysis of the eye-tracking data revealed that participants looked longer at Ds of higher value. This value-based distraction then mediated the reduction of absolute choice accuracy (by leading to more choices of D and by delaying decisions beyond the time limit of 1.5 sec; **Fig. 2c**). Across participants, the distraction of eye movements that was induced by higher values of D correlated with the performance reduction that was induced by higher values of D (**Fig. 2d**). Finally, the patterns of relative and absolute choice accuracy in the novel trials combined over Experiments 1, 2A, and 3 are best accounted for by value-based attentional capture (**Fig. 2e**; compare with predictions shown in **Fig. 1d**). Altogether, these results strongly support value-based attentional capture as the underlying mechanism of sub-optimality in the Chau2014 paradigm.

In Experiment 4 ($n_4 = 44$), we omitted the novel trials to make the task even more similar to Chau2014. Again, we found the value of D to impair absolute choice accuracy, and to increase both D-errors and RTs. Again, we did not find Chau2014's effect on relative choice accuracy (**Table 1**). Finally, an analysis of the original Chau2014 data revealed that D's value increased D-errors and RTs (**Supplementary Table 2**). Hence, even the original dataset supports the value-based attentional capture account.

In conclusion, in none of our four experiments we obtained the significantly negative influence of D's value on relative choice probability (and thus the independence violation) reported in Chau2014. Given that one of our experiments was nearly identical and three experiments were extremely similar to the original study, the absence of a negative HV-D effect on relative choice accuracy in all of our experiments severely challenges the robustness of this effect. In the Supplementary Information, we provide additional results targeting this

low robustness. This includes a sophisticated test[9] of the detectability of the effect in the original study (**Supplementary Fig. 3**) and a Bayes factor analysis for assessing the evidence in favor of the null hypothesis (**Supplementary Fig. 4**). In addition, we provide two analyses of the Chau2014 dataset demonstrating that the significance of the HV-D effect in the original data depends on the presence of the interaction term (HV-LV)×(HV-D) in the regression analysis (**Supplementary Table 2** and **Supplementary Fig. 5**), which hints toward a statistical artifact[10,11].

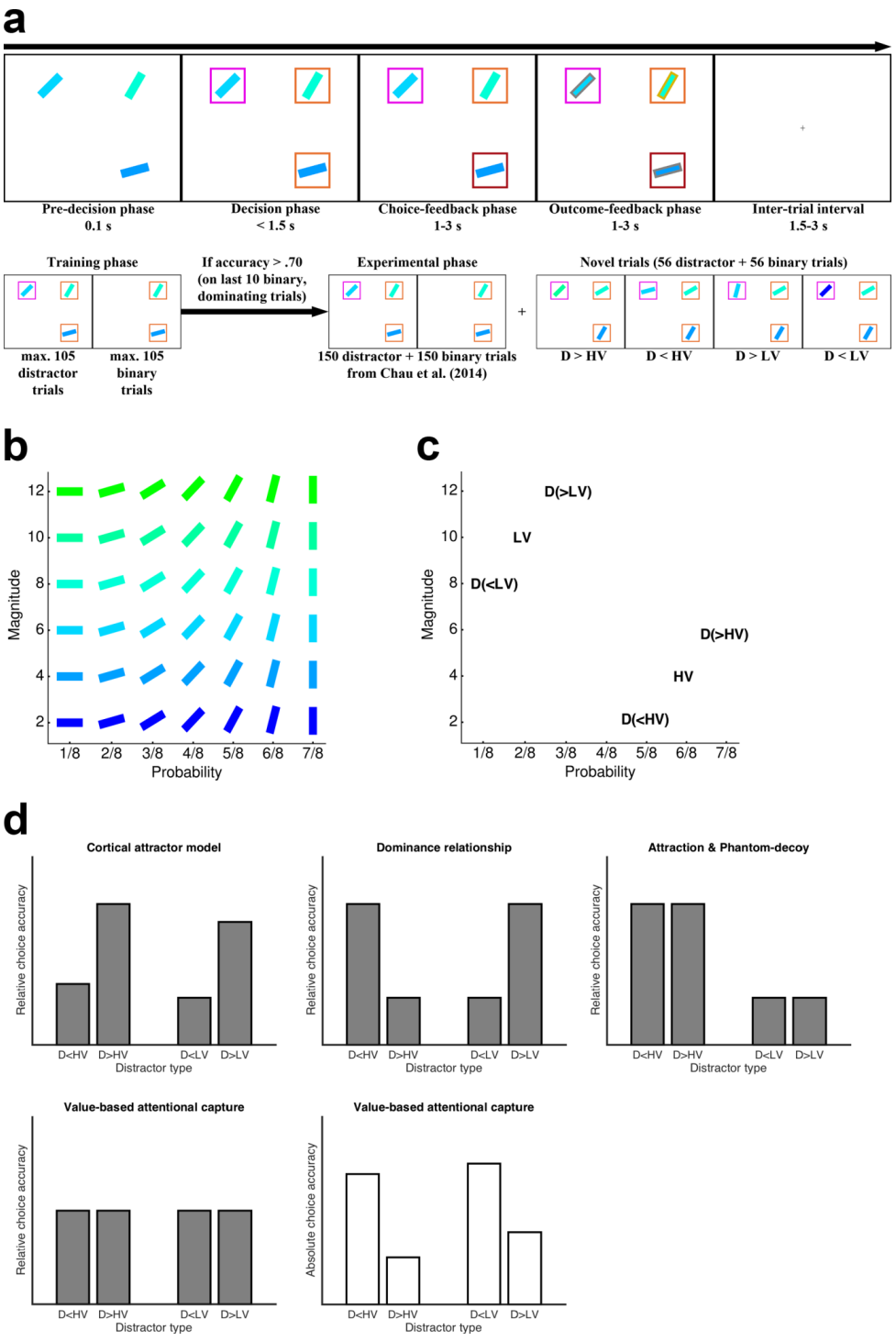
On the other hand, we consistently found that D's value had a deteriorating effect on absolute choice accuracy and decision speed, strongly supporting a value-based attentional capture account. Notably, the paradigm typically used to study value-based attentional capture is very similar to the Chau2014 paradigm: Participants have 1.5 s to identify a target while being distracted by a value-laden stimulus of a specific color[5]. Thus, despite our concerns with respect to the robustness of the effect reported in Chau2014, our work revealed important and novel insights about the influence of value-based attentional capture on value-based decisions and should lead to theory advancement of how people's preferences depend on the choice context.

1. Louie, K., Khaw, M.W. & Glimcher, P.W. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 6139–6144 (2013).
2. Chau, B.K.H., Kolling, N., Hunt, L.T., Walton, M.E. & Rushworth, M.F.S. *Nat. Neurosci.* **17**, 463–470 (2014).
3. Hunt, L.T., Dolan, R.J. & Behrens, T.E.J. *Nat. Neurosci.* **17**, 1613–1622 (2014).
4. Gluth, S., Hotaling, J.M. & Rieskamp, J. *J. Neurosci.* **37**, 371–382 (2017).
5. Anderson, B.A., Laurent, P.A. & Yantis, S. *Proc. Natl. Acad. Sci.* **108**, 10367–10371 (2011).
6. Huber, J., Payne, J.W. & Puto, C. *J. Consum. Res.* **9**, 90–98 (1982).
7. Pratkanis, A.R. & Farquhar, P.H. *Basic Appl. Soc. Psychol.* **13**, 103–122 (1992).
8. Pettibone, J.C. *Judgm. Decis. Mak.* **7**, 513–523 (2012).
9. Simonsohn, U. *Psychol. Sci.* **26**, 559–569 (2015).
10. Gelman, A. *J. Theor. Biol.* **245**, 597–599 (2007).
11. Ai, C. & Norton, E.C. *Econ. Lett.* **80**, 123–129 (2003).

ACKNOWLEDGEMENTS

This work was supported by the Swiss National Science Foundation (SNSF Grant 100014_153616 to S.G. and J.R.). We thank Bolton Chau and Matthew Rushworth for providing us with the behavioral data from the original study and for helpful and constructive discussions and comments.

7


a
Pre-decision phase
0.1 s
Decision phase
< 1.5 s
Choice-feedback phase
1-3 s
Outcome-feedback phase
1-3 s
Inter-trial interval
1.5-3 s
Training phase
max. 105 distractor trials
max. 105 binary trials
If accuracy > .70
(on last 10 binary, dominating trials)
Experimental phase
150 distractor + 150 binary trials from Chau et al. (2014)
Novel trials (56 distractor + 56 binary trials)
D > HV
D < HV
D > LV
D < LV
b
Magnitude
Probability
1/8
2/8
3/8
4/8
5/8
6/8
7/8
c
D(>LV)
LV
D(<LV)
D(>HV)
HV
D(<HV)
d
Cortical attractor model
Dominance relationship
Attraction & Phantom-decoy
Value-based attentional capture
Value-based attentional capture
Relative choice accuracy
Absolute choice accuracy
D<HV
D>HV
D<LV
D>LV
Distractor type

8

Figure 1 Study design and predictions for novel trials. (**a**) Example trial (upper panel) and general workflow (lower panel) of the Chau2014 paradigm as used in our four experiments (variations in specific experiments are outlined in the main text and **Supplementary Methods**). (**b**) Stimulus matrix showing one (of four possible) associations of color and orientation of rectangles with reward probability and magnitude. (**c**) Example of a set of four novel trials (HV and LV are kept constant across these four trials, but D varies). (**d**) Qualitative predictions of choice accuracy in the novel trials for the biophysical cortical attractor model proposed by Chau2014, our initial hypothesis, a combination of attraction and phantom-decoy effects, and value-based attentional capture. The predictions vary with respect to the factors Similarity (i.e., whether D is more similar to HV or to LV) and/or Dominance (i.e., whether D is better or worse than HV/LV). In contrast to the other models, value-based attentional capture does not predict any influence of D's value on relative choice accuracy (gray bars) but a detrimental effect on absolute choice accuracy (white bars).

9

Table 1 Regression analyses of relative and absolute choice accuracy, D-errors, and response times for all experiments.

	Relative choice accuracy				
	Exp. 1	Exp. 2A	Exp. 3	Exp. 4	Combined
HV-LV	**0.426*** **	**0.498*** **	**0.444*** **	**0.551*** **	**0.489*** **
HV+LV	**-0.097***	**-0.172****	**-0.170****	**-0.153*** **	**-0.146*** **
HV-D	-0.027	0.100	-0.096	-0.065	-0.028
(HV-LV)×(HV-D)	0.075	-0.092	**0.141***	0.065	0.050^{+}
D present	**-0.078***	**-0.086***	**-0.068***	**-0.104*** **	**-0.087*** **

	Absolute choice accuracy				
	Exp. 1	Exp. 2A	Exp. 3	Exp. 4	Combined
HV-LV	**0.371*** **	**0.401*** **	**0.376*** **	**0.448*** **	**0.405*** **
HV+LV	-0.057	**-0.139****	**-0.127***	**-0.093*** **	**-0.100*** **
HV-D	0.096	**0.198****	0.048	0.070^{+}	**0.098*** **
(HV-LV)×(HV-D)	0.076	-0.029	**0.180****	**0.089***	**0.079****
D present	**-0.178*** **	**-0.179*** **	**-0.160*** **	**-0.229*** **	**-0.193*** **

	Frequency of choosing D				
	Exp. 1	Exp. 2A	Exp. 3	Exp. 4	Combined
(Value of) D	**0.463*** **(1)	**0.434*** **	**0.412*** **	**0.396*** **(1)	**0.424*** **(2)

	Response times (of HV and LV choices)				
	Exp. 1	Exp. 2A	Exp. 3	Exp. 4	Combined
HV-LV	**-13.2*** **	**-12.8*** **	**-12.0*** **	**-19.6*** **	**-15.2*** **
HV+LV	**-42.9*** **	**-41.8*** **	**-40.4*** **	**-45.6*** **	**-43.2*** **
(Value of) D	**16.5*** **	**14.9*** **	**9.8****	**12.2*** **	**13.4*** **
D present	**64.8*** **	**60.1*** **	**60.1*** **	**69.6*** **	**64.7*** **

Note. The values represent average standardized regressions coefficients (intercepts omitted). Only the two sets of 150 trials each (with D present or absent) from Chau2014 but not the novel trials were included in these analyses. The gray shading highlights the crucial HV-D

10

predictor on relative choice accuracy, for which Chau2014 found a significantly negative regression coefficient (but we did not in any of our four experiments). For the analysis of the frequency of choosing D, values in parentheses indicate the number of participants that never chose D and were thus excluded from this analysis. Regression analyses for response times include (and thus control for) additional factors influencing response times. Results for Experiment 2, Group B are not reported here, as the task deviated from Chau2014 with respect to deliberation time (see main text). $^{+}P < .1$, $^{*}P < .05$, $^{**}P < .01$, $^{***}P < .001$.

11

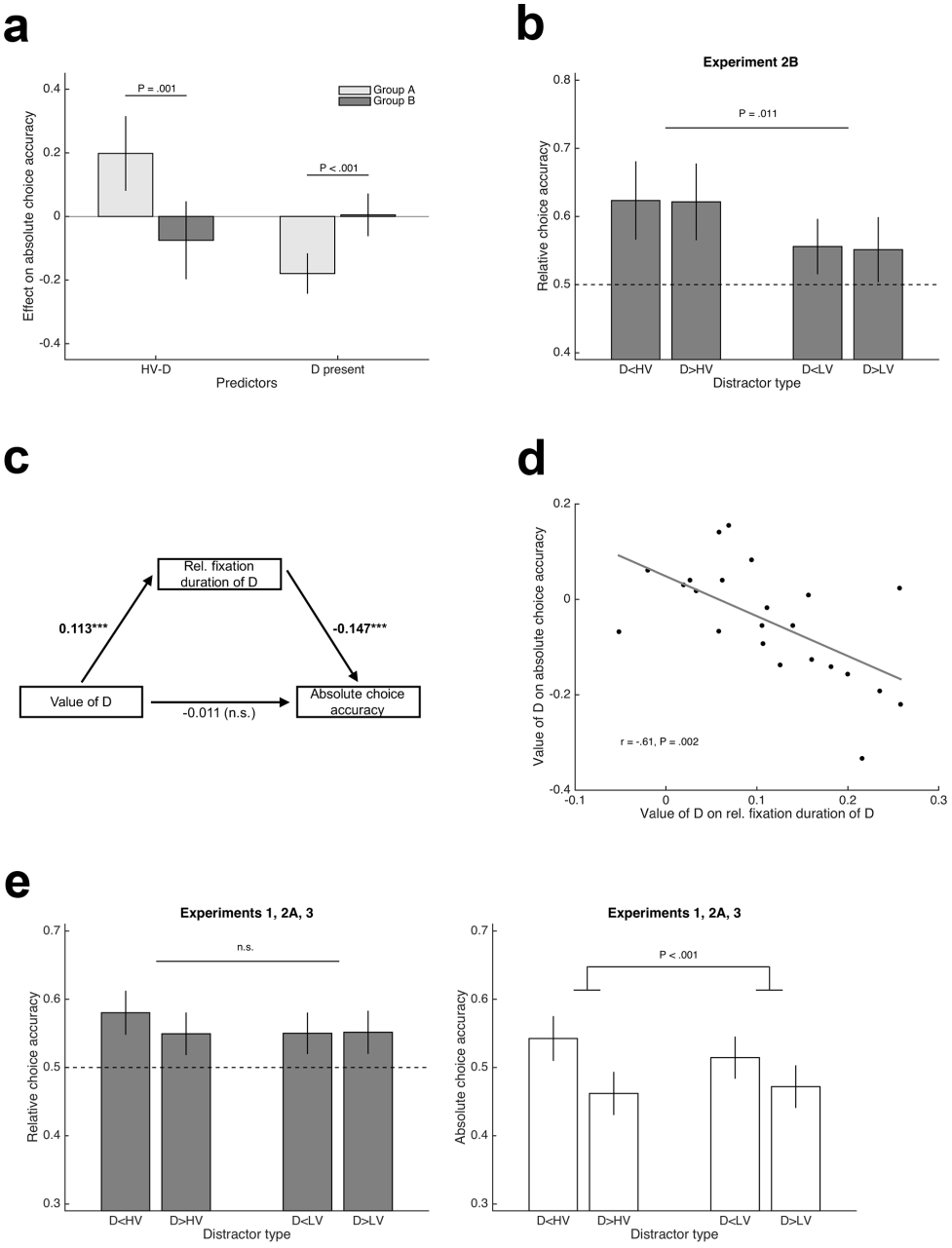


Figure 2 Results of Experiments 2 (**a, b**), 3 (**c, d**), and 1-3 combined (**e**). (**a**) Group comparison of regression coefficients reflecting the influence of D's value and presence on absolute choice accuracy. D did not distract decisions in Group B (with extended deliberation time of 6 s). (**b**) The analysis of Group B's relative choice accuracy in novel trials indicates a pattern of independence violations that are best explained by combined attraction and phantom-decoy effects (cf. **Fig. 1d** and **Supplementary Methods**; note that the comparatively low performance in these novel trials reflects the high difficulty in these trials

12

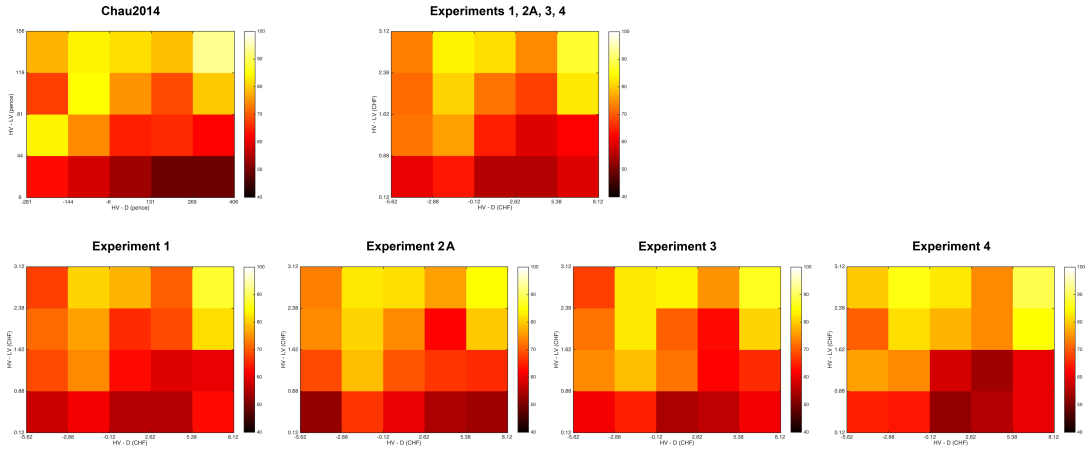
due to similar expected values of HV and LV; see **Fig. 1c**). (**c**) A path analysis of the eye-tracking data shows that participants looked more at high-value Ds, which in turn reduced their choice accuracy (values represent standardized coefficients; ***$P < .001$). (**d**) Consistent with this, participants whose attention on D was driven more strongly by D's value also exhibited a stronger (negative) influence of D's value on their choice accuracy. (**e**) Relative (left panel) and absolute (right panel) choice accuracy in novel trials across all experiments with short deliberation. Consistent with value-based attentional capture, D's value does not decrease relative but only absolute choice accuracy (cf. **Fig 1d**). Error bars represent 95% confidence intervals.

SUPPLEMENTARY INFORMATION

Value-based attentional capture affects multi-alternative decision making

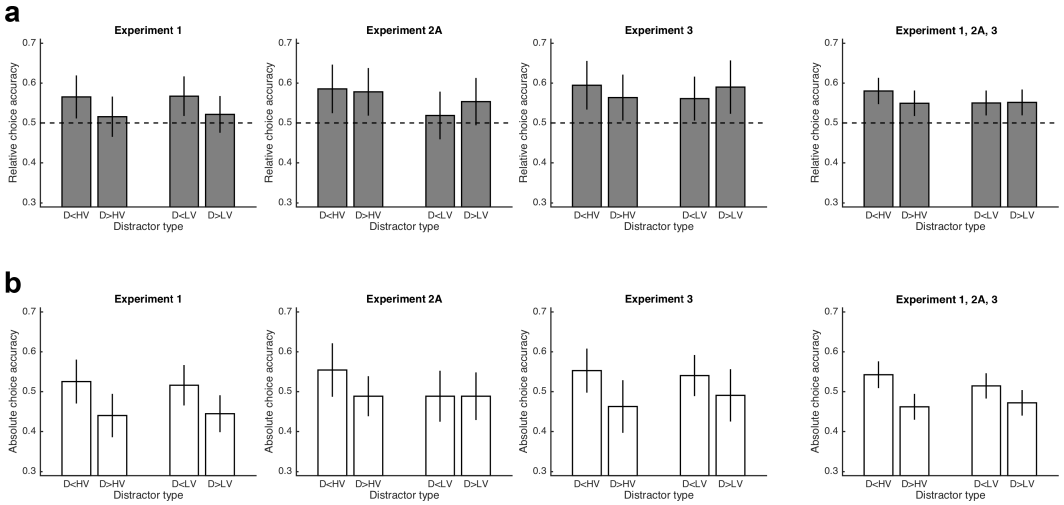
Sebastian Gluth*, Mikhail S. Spektor* & Jörg Rieskamp

*Equal contribution

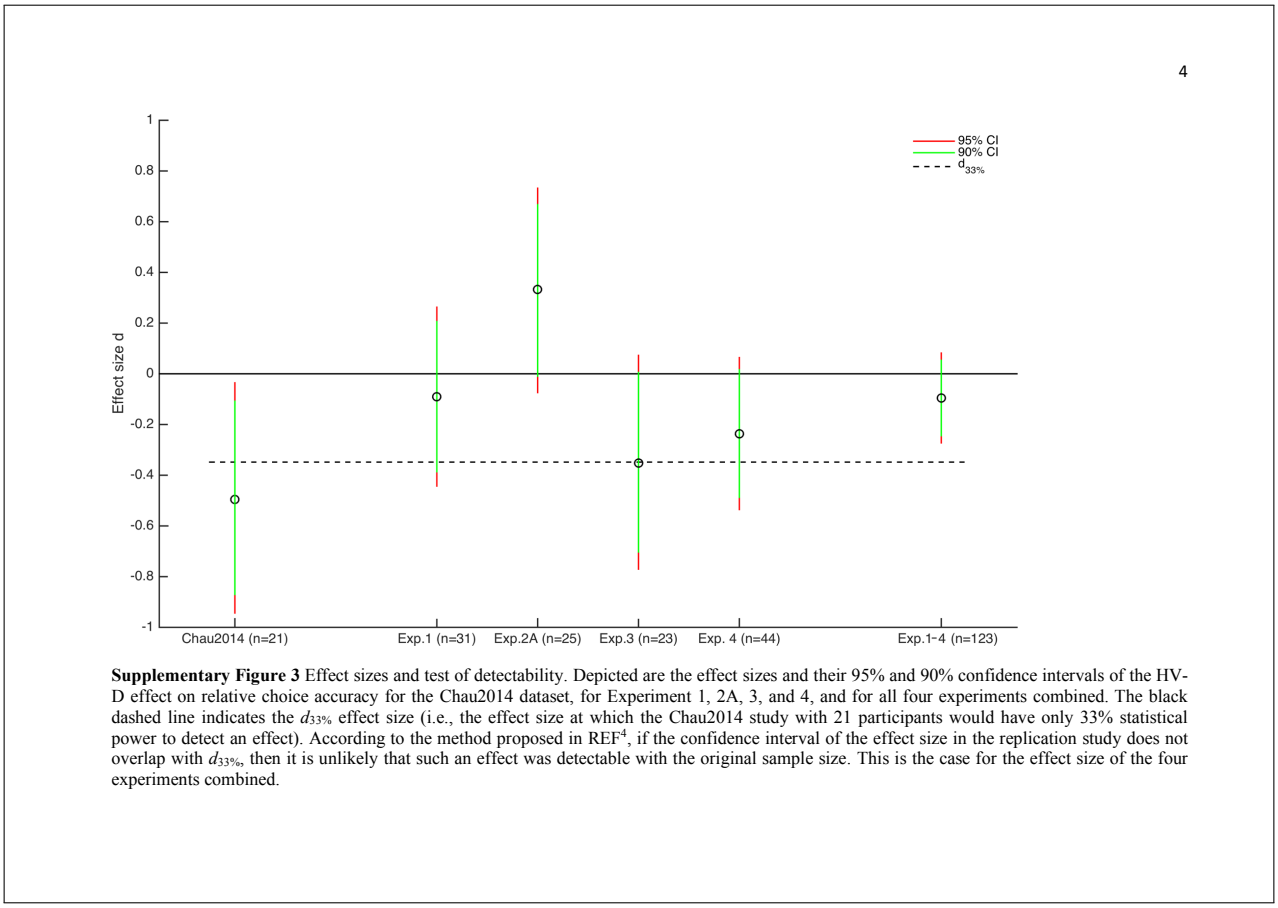


Supplementary Figure 1 Comparison of relative accuracy with Chau2014[1] dataset. Depicted are the "accuracy matrices" (i.e., relative choice accuracy as a function of HV-D and HV-LV; see Fig. 2c in Chau2014, p. 465) for the original dataset, for all our experiments combined, and for each of our four experiments separately. There is no evidence for overall performance differences with respect to relative choice accuracy. However, the decrease in accuracy as a function of HV-D (i.e., from left to right) reported in Chau2014 is absent in all our experiments.

3

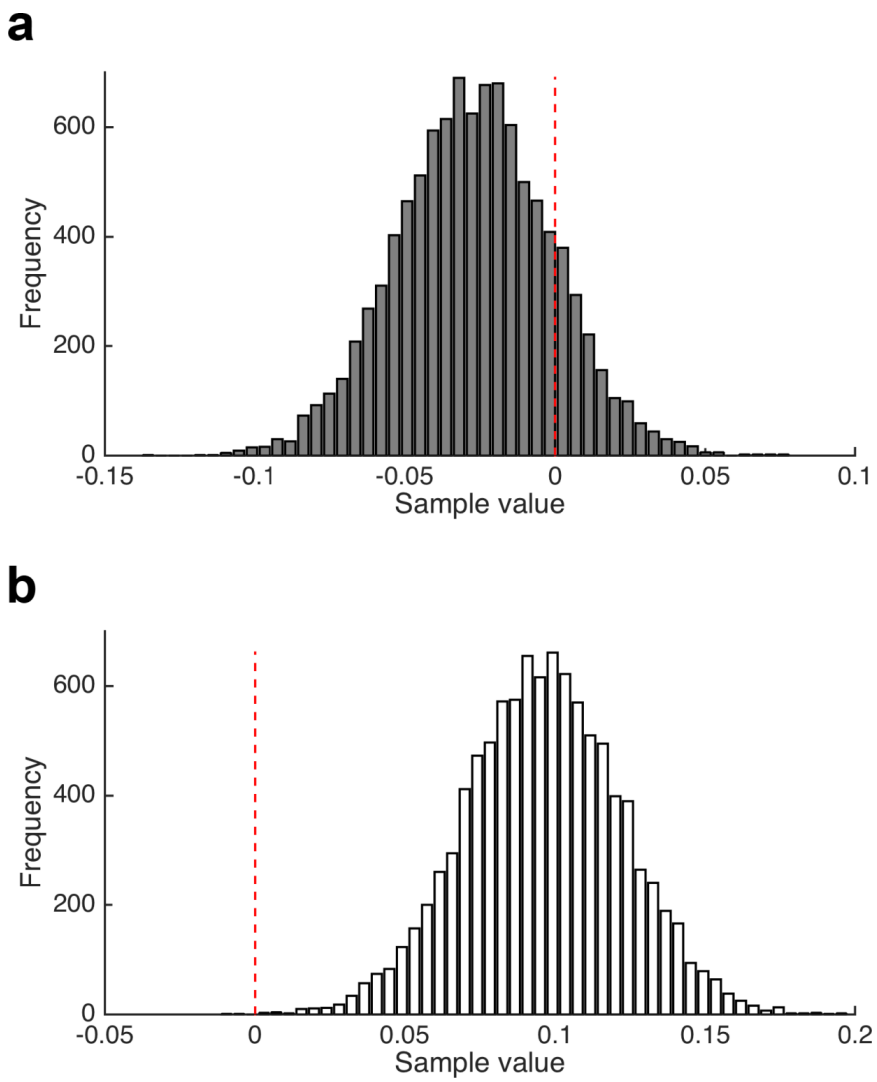


Supplementary Figure 2 Relative and absolute choice accuracy in novel trials in experiments with short deliberation time (as in Chau2014). (**a**) Relative choice accuracy in novel trials separated by the position of D relative to HV and LV (cf. **Fig. 1** and **Supplementary Methods**) for Experiments 1 to 3 and combined. In Experiment 1, there is a significant main effect of Dominance that is opposite to the effect reported in Chau2014 (i.e., relative choice accuracy is higher when D is dominated by HV or LV; $F_{1,30} = 5.85$, $P = .022$). This effect would be consistent with a divisive normalization account[2]. However, the effect was not found in the other experiments (combined: $P > .28$). (**b**) The same analyses for absolute choice accuracy. Here, the main effect of Dominance is robust (combined: $F_{1,78} = 17.73$, $P < .001$). In line with a value-based attentional capture account[3], this effect is driven by the fact that participants chose D more often when it had a higher value. Note that the comparatively low (but still better than chance) performance in the novel trials is due to the nearly identical expected values of HV and LV in these trials (cf. **Fig. 1c**).

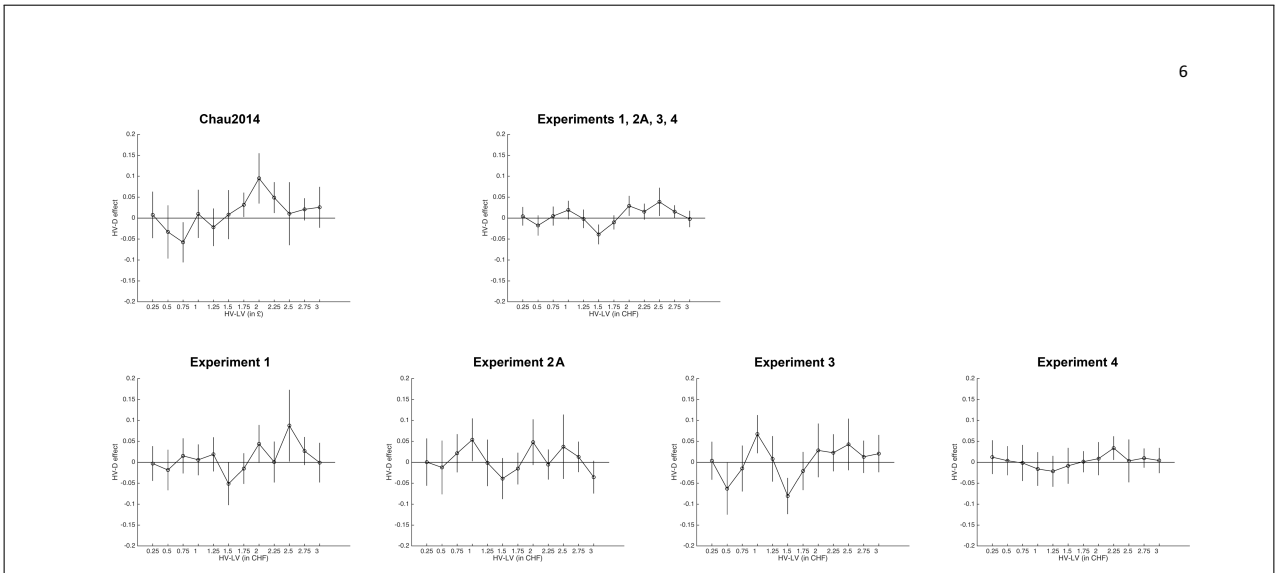


Supplementary Figure 3 Effect sizes and test of detectability. Depicted are the effect sizes and their 95% and 90% confidence intervals of the HV-D effect on relative choice accuracy for the Chau2014 dataset, for Experiment 1, 2A, 3, and 4, and for all four experiments combined. The black dashed line indicates the $d_{33\%}$ effect size (i.e., the effect size at which the Chau2014 study with 21 participants would have only 33% statistical power to detect an effect). According to the method proposed in REF[4], if the confidence interval of the effect size in the replication study does not overlap with $d_{33\%}$, then it is unlikely that such an effect was detectable with the original sample size. This is the case for the effect size of the four experiments combined.

5



Supplementary Figure 4 Bayesian analysis of the HV-D effect on relative and absolute choice accuracy. (**a**) Histogram of posterior samples from a Bayesian *t*-test (using the default settings of the *R* package *BayesFactor*) for the HV-D regression coefficients on relative choice accuracy. The analysis includes the coefficients from all 123 participants that conducted the task with short deliberation time (as in Chau2014). As can be seen, the posterior distribution largely overlaps with 0 (the Bayes Factor in favor of the null hypothesis relative to the alternative hypothesis is 5.78, which is also seen as *substantial*[5] or *positive*[6] evidence for the null hypothesis). (**b**) The same analysis for absolute choice accuracy. Here, there is almost no overlap with 0 (the Bayes factor in favor of the alternative hypothesis relative to the null hypothesis is 78.63, which is seen as *strong*[5,6] evidence for the alternative hypothesis).



Supplementary Figure 5 Analysis of the HV-D effect on relative choice accuracy for different levels of HV-LV and without (HV-LV)×(HV-D). Our analysis of the Chau2014 dataset suggested that the HV-D effect was mainly driven by the presence of the interaction term (HV-LV)×(HV-D) in the regression analysis (see **Supplementary Table 2**). Therefore, we analyzed the effect of HV-D for each of the 12 HV-LV differences separately (this allows to and requires taking the predictor variables HV-LV and the interaction between HV-LV and HV-D out of the logistic regression). Strikingly, a robust HV-D effect was obtained in none of our own experiments but also not in the Chau2014 dataset. For the latter, the HV-D effect is significantly negative in only 1 of 12 levels of HV-LV but significantly positive in three other levels. Merging over all levels, the effect of HV-D shows a trend in the positive direction (t_{20} = 2.04, P = .055, d = 0.45). Even when restricting the analysis to the 6 most difficult levels, there is no significantly negative HV-D effect (t_{20} = -1.25, P = .227, d = -0.27). This confirms the notion that the inclusion of the interaction term is the main reason for the HV-D effect being significantly negative in the Chau2014 dataset.

Supplementary Table 1 Comparison of performance with Chau2014 dataset.

			Mean			
	Chau2014	Experiment 1	Experiment 2A	Experiment 3	Experiment 4	Combined
Relative; all trials	.715	.688 (.124)	.709 (.721)	.712 (.904)	.713 (.938)	.706 (.543)
Absolute; all trials	.667	.628 (.072^{+})	.651 (.406)	.661 (.775)	.645 (.250)	.645 (.199)
Relative; trials without D	.727	.703 (.180)	.726 (.971)	.726 (.958)	.732 (.783)	.722 (.768)
Absolute; trials without D	.699	.668 (.132)	.690 (.636)	.697 (.910)	.694 (.773)	.687 (.471)
Relative; trials with D	.701	.670 (.147)	.689 (.552)	.698 (.848)	.693 (.608)	.687 (.398)
Absolute; trials with D	.635	.587 (.091^{+})	.611 (.330)	.626 (.700)	.596 (.080^{+})	.602 (.129)
			Variance (in SD)			
Relative; all trials	.052	.066 (.267)	.062 (.440)	.067 (.266)	.059 (.533)	.063 (.320)
Absolute; all trials	.065	.082 (.267)	.067 (.865)	.069 (.788)	.074 (.517)	.074 (.499)
Relative; trials without D	.062	.062 (.970)	.063 (.919)	.076 (.365)	.067 (.731)	.067 (.719)
Absolute; trials without D	.072	.073 (.970)	.059 (.365)	.077 (.780)	.076 (.826)	.072 (.950)
Relative; trials with D	.058	.087 (.060^{+})	.075 (.232)	.067 (.501)	.062 (.752)	.073 (.240)
Absolute; trials with D	.074	.111 (.060^{+})	.087 (.458)	.081 (.664)	.084 (.518)	.092 (.260)

Note. "Relative" refers to relative choice accuracy; "Absolute" refers to absolute choice accuracy. Values in parentheses refer to *P*-values of two-sample *t*-tests / two-sample *F*-tests for comparing means / variances between the Chau2014 dataset and our datasets. $^{+}P < .1$.

8

Supplementary Table 2 Analyses of the original Chau2014 dataset.

Relative choice accuracy without (HV-LV)×(HV-D)

	Coefficient	*P*-value
HV-LV	**0.573**	**< .001**
HV+LV	**-0.180**	**.003**
HV-D	0.022	.510
D present	-0.066	.068

Frequency of choosing D

	Coefficient	P-value
(Value of) D	**0.371**	**< .001**

Response times (of HV and LV choices)

	Coefficient	P-value
HV-LV	**-19.54**	**< .001**
HV+LV	**-42.19**	**< .001**
(Value of) D	**6.78**	**.008**
D present	**77.18**	**< .001**

Note. Values in the middle column represent average standardized regression coefficients (intercepts omitted). The gray shading highlights the HV-D predictor on relative choice accuracy. As can be seen, the coefficient is not significantly negative anymore when the interaction term (HV-LV)×(HV-D) is removed from the regression analysis (see also **Supplementary Figure 5**). On the other hand, D's value leads to more frequent choices of D and to slower responses when choosing HV or LV, which is in line with the data of our experiments and a value-based attentional capture account.

SUPPLEMENTARY METHODS

Participants. Thirty-one participants (21 female, age: 20-47, *M*= 27.71, *SD* = 6.59, 29 right-handed) completed Experiment 1. A total of 51 participants signed up for Experiment 2. Due to computer crashes, the data of two participants (one from Group A and one from Group B) were incomplete and excluded from the analyses, resulting in a final sample of 49 participants, 25 were in Group A (13 female, age: 20-46, *M* = 26.88, *SD* = 6.62, 24 right-handed) and 24 were in Group B (11 female, age: 19-35, *M* = 23.96, *SD* = 3.43, 20 right-handed). Thirty participants signed up for the eye-tracking Experiment 3. One participant was excluded for not passing the training-phase criterion, and additional six participants were excluded due to incompatibility with the eye-tracking device (for further details see below), resulting in a final sample of 23 participants (14 female, age: 18-54, *M* = 25.70, *SD* = 8.66, 19 right-handed). Forty-seven participants signed up for Experiment 4. Due to failing the training-phase criterion, three participants were excluded, resulting in a final sample of 44 participants (36 female, age: 18-46, *M* = 23.70, *SD* = 5.74, 40 right-handed). Data of all participants included in the final samples are made publicly available on the Open Science Framework at https://osf.io/8r4fh/ (upon publication of the study).

Paradigm. The paradigm was very similar for all four experiments. Participants repeatedly chose between either two (binary trials) or three (distractor trials) two-outcome lotteries (gambles), each yielding an outcome of magnitude X in Swiss francs (CHF) with probability *p*, or 0 otherwise. The gambles were represented by colored rectangles, each shown in a random quadrant of the screen, whereby the rectangles' colors represented outcomes X and the angles represented the probabilities *p*. Outcomes ranged from CHF 2 to CHF 12 in steps of CHF 2 and were represented by colors ranging from either green to blue or blue to green. Probabilities ranged from 1/8 to 7/8 in steps of 1/8 and were represented by orientation angles

ranging from 0° to 90° or from 90° to 0° in steps of 15° (see **Fig. 1b** in the main text for all colors and orientations). Associations between colors/outcomes and orientations/probabilities were counter-balanced between participants. In binary trials, participants saw the two options for 100 ms before orange frames appeared around each option (pre-decision phase). After the frames appeared, participants had up to 1.5 s to make a choice by pressing 7, 9, 1, or 3 on the numeric keypad for upper left, upper right, lower left, or lower right quadrant, respectively (participants belonging to Group B of Experiment 2 had 6 s after appearance of the frames to decide). Distractor trials were similar to binary trials: All of the options were presented for 100 ms before frames appeared around them. Contrary to the binary trials, one of the options had a magenta frame (the distractor), signaling that it could not be chosen. Choosing the distractor resulted in a screen telling that the option was not available after which a new trial began. Similarly, choosing an empty quadrant resulted in a screen showing that the quadrant was empty after which a new trial began.

If a valid choice was registered, the trial continued with a choice-feedback phase for 1-3 s, in which the chosen option was highlighted in a dark red color. To make sure that participants pay attention to all available options, with a probability of 15% participants had to complete a "match" trial before the choice-feedback phase. On match trials, one of the options from the decision phase (including the distractor, if distractor trial) was presented in the middle of the screen. Participants had up to 2 s (6 s in Experiment 2, Group B) to press the key corresponding to the option's quadrant. If correct, participants saw a screen saying "correct" and extra CHF 0.10 were added to the participant's account, otherwise they saw a screen saying "wrong". After every match trial, the trial continued with the choice-feedback phase.

After the choice-feedback phase participants received feedback about the outcomes of the gambles for 1-3 s. The frames' colors changed to grey if the option did not yield a reward (i.e., the outcome was CHF 0) or to golden yellow if the option yielded a reward. Participants

also received feedback about the distractor's outcome on distractor trials. After this outcome-feedback phase, a new trial began with an inter-trial interval of 1.5-3 s in which a fixation cross was shown.

Experimental procedure. After giving informed consent and filling out the demographic questionnaire, participants received detailed instructions about the task and were familiarized with the outcome and probability associations by making six judgments for each dimension in a paired comparison. In all experiments, participants completed a training phase and an experimental phase. In the training phase, participants encountered up to 210 trials, half of which were distractor trials. These trials were randomly generated with the boundary condition that 2/3 of the trials were not dominant (i.e., HV did not have a higher probability *and* a higher outcome than LV) and the rest were dominant. The training phase continued until participants encountered at least 10 dominant binary trials, and chose HV in at least 70% of the last 10 encountered dominant binary trials. The training phase ended when this criterion was reached and the experimental phase began. If the participants finished all 210 training trials without passing the criterion, the experiment ended. As reported above, participants who did not pass the criterion were excluded from the analysis.

The experimental phase consisted of either 412 (Experiment 1, 2, & 3) or 300 (Experiment 4) trials. The 300 trials used in Experiment 4 were shared across all experiments and are those used by Chau2014. In Experiment 4, all trials were presented in exactly the same orders as in Chau2014's experiment, whereas in the other experiments, only the distractor trials were presented in the order provided to us by the authors of Chau2014 with the randomized binary trials interleaved. In addition to these trials, Experiments 1–3 included 56 novel distractor trials and, correspondingly, the 56 binary trials belonging to these novel trials (details are provided below). Throughout the experiment, participants had the opportunity to make four breaks. After completing the experiment, participants received their

show-up fee (CHF 5 for 15 minutes), the average reward of the chosen option (distractor and empty quadrant choices counted as no reward), and the accumulated match bonuses. If participants reached the experimental phase, the experiments took approximately 75 minutes.

Initial hypothesis for the effects reported in Chau2014. Before conducting our experiments, we assumed that the positive relationship between the value of *D* and relative choice accuracy as reported by Chau2014 was a robust effect. To explain the apparent contradiction with the findings of Louie and colleagues[2], we reasoned that the explicit presentation of two attributes (i.e., magnitude X and probability *p* of reward) in the Chau2014 task led people to compare the options on those attributes directly (i.e., a multi-attribute decision between attributes X and *p* instead of the expected values of the options). Importantly, certain attribute-wise comparison processes are known to produce violations of independence, so-called "context effects of preferential choice"[7–11]. More specifically, our initial hypothesis was that individuals can recognize that D is either better or worse than LV and/or HV with respect to each attribute (e.g., D might dominate LV with respect to probability). Critically, since LV is per definition worse than HV, it is more likely that D dominates LV than that it dominates HV on some attribute. However, this is only true as long as D has not very low attribute values (leading to a low expected value overall). Thus, the "dominance relationship" between D and LV/HV may be a useful indicator helping to identify the option with the highest expected value. The positive relationship between D and relative choice accuracy would then be an epiphenomenon of this "dominance relationship" mechanism.

To give an example, let us assume that HV, LV, and D are specified as follows:

HV: $p = 5/8$; X = CHF 8 → EV = CHF 5

LV: $p = 3/8$; X = CHF 4 → EV = CHF 1.5

D: $p = 6/8$; X = CHF 6 → EV = CHF 4.5

In this case, D is superior to HV and LV with respect to probability. With respect to magnitude, however, D is superior to LV but is inferior to HV. Hence, by counting the number of times D is superior to LV and HV on the attributes (i.e., 2 for LV vs. 1 for HV), a decision maker could correctly identify the option with the highest value. Critically, this information is not helpful anymore when we replace D by a distractor D* of lower expected value:

D*: $p = 6/8$; X = CHF 2 → EV = CHF 1.5

In this new case, the distractor is inferior to both HV and LV with respect to magnitude and (still) superior to both with respect to probability (i.e., the count is 1 for LV vs. 1 for HV). Thus, the option with the highest expected value cannot be identified anymore based on the dominance relationship alone. This demonstrates that the high-value D might better support choosing between HV and LV than the low-value D*. Notably, the idea that the relative ranking of options influences decision making has also been suggested by others[12–14].

Besides this *dominance relationship* hypothesis, the paradigm of Chau2014 might also involve other context effects and influence behavior. We considered the *attraction effect*[11,15] and the *phantom-decoy effect*[9,16]. According to the attraction effect, the preference between two options (in our case between HV and LV) can be changed by adding a third option (in our case D) that is similar but clearly inferior to only one of the two options (in our case, if D is similar to HV and worse than it, HV should be preferred; if D is similar to LV and worse than it, LV should be preferred). The phantom-decoy effect predicts that if an option is similar but worse than D (and D is unavailable, as in the Chau2014 task), it is more likely to be chosen. Finally, a combination of attraction and phantom-decoy effects predicts that independent of whether D is worse or better than the similar option, the option that is more similar to D is more likely to be chosen. As described in the following section, we sought to distinguish between all these different context effects and the model proposed by Chau2014 by implementing a novel set of trials.

Importantly, in contrast to the context effects, the Chau2014 model, and the divisive normalization account, value-based attentional capture does not predict any influence of D on relative choice accuracy–but a negative effect on absolute choice accuracy (see **Fig. 1d** in the main text).

The novel trial set and predictions of different models and context effects. In the novel set of trials used to dissociate predictions from various models and context effects, the HV and LV options were arranged such that i) HV was superior to LV with respect to one attribute (magnitude or probability) but ii) inferior with respect to the other attribute, and iii) D could be placed such that it either fully dominated HV/LV or it was fully dominated by HV/LV (for an example, see **Fig. 1c** in the main text). In the Chau2014 task, there are 14 possible combinations of HV and LV that fulfill these criteria. For each of these 14 combinations, D was placed directly "above" or "below" HV or LV resulting in 4 trials per combination and 56 novel trials in total (we also added 56 binary trials without *D*, so that we had 112 trials more than the original study).

The (qualitative) predictions of the models and context effects with respect to these novel trials are outlined in **Figure 1d**. Chau2014's biophysical cortical attractor model predicts a positive effect of the value of D on relative choice accuracy. Thus, the performance should be higher in trials with dominant distractors (i.e., D>HV and D>LV; note that the divisive normalization model by Louie and colleagues[2] predicts higher accuracy for low-value Ds, or in other words, it predicts the opposite of Chau2014's model). Our initial dominance relationship hypothesis predicts higher choice accuracy if HV dominates D (i.e., D<HV), or if LV is dominated by D (i.e., D>LV), because in these cases HV is better than D on more attributes than LV. The combination of attraction and phantom-decoy effects predicts more accurate choices when D is dominated by HV (i.e., D<HV) or dominates HV (i.e., D>HV), or in other words, when D is more similar to HV than to LV. Importantly, all these predictions

refer to relative choice accuracy (for which we do not find any robust effects in the Chau2014 task with short deliberation time; **Supplementary Fig. 2**). On the contrary, value-based attentional capture predicts no effect on relative choice accuracy, but a reduction of absolute choice accuracy when D has a high value (i.e., D>HV and D>LV).

Note that the different predictions can also be formulated in terms of main effects and interactions of an ANOVA with the factors Dominance (D dominates or is dominated by HV/LV) and Similarity (D is more similar to HV or to LV). Within this ANOVA, the cortical attractor model predicts a main effect of Dominance (the divisive normalization model also predicts this but in the opposite direction). The dominance relationship hypothesis predicts an interaction effect of Dominance and Similarity. The combined attraction/phantom-decoy effect predicts a main effect of Similarity. Value-based attentional capture predicts the same main effect of Dominance as the divisive normalization model but on absolute (not relative) choice accuracy.

Behavioral data analysis. In each trial, there was always a higher-valued option HV and a lower-valued option LV. We used two different dependent measures, *relative* and *absolute* choice accuracy. Relative choice accuracy refers to the proportion of HV choices among choice of HV and LV only, whereas absolute choice accuracy refers to the proportion of HV choices among all choices (including missed responses due to the time limit). For each of the two dependent variables, we estimated intra-individual logistic regressions and tested the regression coefficients between subjects against 0 using a two-sided one-sample *t*-test with an α level of .05. We used the set of predictor variables reported in Chau2014, which consisted of the difference in expected value (EV) of the two available options, HV-LV, the sum of their EVs, HV+LV, the EV difference between HV and D, HV-D, the interaction between HV-LV and HV-D, (HV-LV)×(HV-D), and whether it was a binary or distractor trial, D present. In the binary trials, the predictors HV-D and (HV-LV)×(HV-D) were kept constant

(i.e., replaced by the mean values in the distractor trials in the regression analysis). All predictor variables were z-transformed to obtain standardized regression coefficients. In addition, we analyzed the influence of the value of D on the tendency to choose D (logistic regression) and on the mean RT of HV and LV choices (linear regression). The RT analysis included additional predictor variables with a significant influence on RT (i.e., HV-LV, HV+LV, D present). Because participants received feedback after each decision, we tested whether improvements of choice accuracy over time affected the results by re-analyzing all regressions with an additional predictor variable that coded for the (standardized) trial number. Although we found a modest learning effect on absolute choice accuracy ($t_{122} = 2.12$, $P = .036$), none of the reported results were affected by including this control variable. Note that only the 300 trials that were also used in the Chau2014 paradigm (but not our 112 novel trials) were included in these regression analyses. The novel trial sets used to dissociate different model predictions were analyzed by a 2 (Dominance) x 2 (Similarity) ANOVA (for details see above).

Eye-tracking procedure and analysis. Experiment 3 was conducted while participants' gaze positions were recorded using an SMI RED500 eye-tracking device. The experimental procedure was adapted to make it suitable for an eye-tracking experiment. Participants completed the experiment on a 47.38x29.61 cm screen (22" screen diagonal) with a resolution of 1680x1050 pixels. During the inter-trial interval, participants were instructed to look at the fixation cross and the random duration of the inter-trial interval was removed. Instead, there was a real-time circular area of interest (AOI) with a diameter of 200 pixels around the fixation cross. Participants' gazes had to (continuously) stay within this AOI for 1 s for the trial to begin. This was done to make sure that participants were indeed looking at the fixation cross and to check the calibration at every trial. If this criterion was not reached within 12 s, the eye tracker was re-calibrated. This procedure was explained to the participants by the

experimenter. In case these re-calibrations happened too frequently (e.g., three times in a row within the same trial, or at least three times in ten trials), the sampling frequency was reduced from the initial 500Hz to 250Hz. If this frequency correction happened repeatedly until the lowest possible frequency of 60Hz was reached, the experiment was aborted and participants received their show-up fee and decision-based bonuses accumulated until then. The first calibration took place just before the training phase and the eye tracker was re-calibrated after each of the four breaks. This experiment took approximately 90 minutes to complete.

The raw gaze positions were re-coded into events (fixations, saccades, and blinks) in SMI's BeGaze2 software package using the high-speed detection algorithm and default values. AOIs were defined around the positions where the frames of the options were, and all fixations inside the frame were counted towards the option within that quadrant. Fixations at empty quadrants as well as all fixations outside of the pre-defined AOIs were counted as *empty gazes*. Fixations within a trial were collapsed and summed to form the dependent variables *relative fixation duration* and *number of fixations*. We report results based on the relative fixation duration (i.e., the sum of the duration of all fixations on a specific quadrant divided by the sum of the duration of all fixations on any quadrant). Note that this measure is highly correlated with the number of fixations, which yielded similar results.

Effect size estimation, power analysis, and test of detectability. According to our re-analysis of the Chau2014 dataset, the effect size Cohen's *d* of the negative regression coefficient of HV-D on relative choice accuracy reported in the original work is -0.495. When using this effect size, the statistical power (i.e., the probability of finding a significant HV-D effect given a true HV-D effect with the assumed effect size) for our Experiments 1, 2A, 3, and 4 with sample sizes of 31, 25, 23, and 44 participants is .79, .66, .62, and .89, respectively (determined with the *pwr.t.test* function of the R-package *pwr*, with an alpha-level of 5% for a two-sided one-sample *t*-test). The power for testing the effect with data from the entire sample

of 123 participants is >.999. The probability of obtaining not a single significant effect in any of the four studies, which is the result of our four studies, is .003.

However, these estimations depend on the published effect size, and it is well known that such effect sizes tend to be inflated due to publication bias[4,17,18]. Accordingly, we also conducted a recently proposed *test of detectability* that is independent of the reported effect size of the original study[4]. This approach uses only the sample size (and the statistical design) of the original study to specify the (hypothetical) effect size that would have given the study only 33% statistical power, which can be regarded as undoubtedly insufficient. If the effect size of the study that attempts to reproduce the effect of the original study is significantly below this $d_{33\%}$ threshold, then one can conclude that the studied effect is not large enough to have been detectable with the original sample size. The results of this test for the HV-D effect on relative choice accuracy show that the 90% and 95% confidence intervals of the effect size for our Experiments 1, 2A, 3, and 4 combined are closer to 0 than the $d_{33\%}$ threshold for the Chau2014 sample size of 21 participants (**Supplementary Fig. 3**). Moreover, given the effect size of all our experiments combined (and its 95% confidence interval), the statistical power of the original study was most likely not higher than 22%.

Together with the analysis of the Chau2014 dataset, which suggest that the HV-D effect may only be a byproduct of the inclusion of the interaction between HV-LV and HV-D (see **Supplementary Table 2** and **Supplementary Fig. 5**), these results strongly suggest that only a negligibly small or no HV-D effect exists, or that this effect is only detectable under extremely narrow conditions (which, in our view, would then render this effect practically meaningless).

REFERENCES

1. Chau, B.K.H., Kolling, N., Hunt, L.T., Walton, M.E. & Rushworth, M.F.S. *Nat. Neurosci.* **17**, 463–470 (2014).
2. Louie, K., Khaw, M.W. & Glimcher, P.W. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 6139–6144 (2013).
3. Anderson, B.A., Laurent, P.A. & Yantis, S. *Proc. Natl. Acad. Sci.* **108**, 10367–10371 (2011).
4. Simonsohn, U. *Psychol. Sci.* **26**, 559–569 (2015).
5. Jeffreys, H. (Oxford University Press: Oxford, UK, 1961).
6. Kass, R.E. & Raftery, A.E. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
7. Roe, R.M., Busemeyer, J.R. & Townsend, J.T. *Psychol. Rev.* **108**, 370–392 (2001).
8. Usher, M. & McClelland, J.L. *Psychol. Rev.* **111**, 757–769 (2004).
9. Pettibone, J.C. & Wedell, D.H. *J. Behav. Decis. Mak.* **20**, 323–341 (2007).
10. Trueblood, J.S., Brown, S.D. & Heathcote, A. *Psychol. Rev.* **121**, 179–205 (2014).
11. Gluth, S., Hotaling, J.M. & Rieskamp, J. *J. Neurosci.* **37**, 371–382 (2017).
12. Stewart, N., Chater, N. & Brown, G.D.A. *Cognit. Psychol.* **53**, 1–26 (2006).
13. Tsetsos, K. et al. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 3102–3107 (2016).
14. Howes, A., Warren, P.A., Farmer, G., El-Deredy, W. & Lewis, R.L. *Psychol. Rev.* **123**, 368–391 (2016).
15. Huber, J., Payne, J.W. & Puto, C. *J. Consum. Res.* **9**, 90–98 (1982).
16. Pratkanis, A.R. & Farquhar, P.H. *Basic Appl. Soc. Psychol.* **13**, 103–122 (1992).
17. Button, K.S. et al. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
18. Open Science Collaboration *Science* **349**, aac4716-aac4716 (2015).