The role of visual attention in preferential choice

Model-based analyses of choice and eye movement data

Dissertation

Zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften
(Dr. rer. nat.)

am Fachbereich Erziehungswissenschaft und Psychologie
der Freien Universität Berlin

Vorgelegt von
Felix Molter
M.Sc. Social, Cognitive and Affective Neuroscience
B.Sc. Psychologie

Berlin, 2021

Erstgutachter: Prof. Dr. Hauke R. Heekeren

Zweitgutachterin: Prof. Dr. Soyoung Q Park

Drittgutachter: Prof. Dr. Sebastian Gluth

Datum der Disputation: 18. Februar 2022

Contents

Acknowledgements

I consider myself lucky to have received support from a number of people along the way leading to this thesis.

First and foremost, I wish to thank my supervisors Peter Mohr and Hauke Heekeren. Peter has supported me throughout the years to an extent I consider exceptional, always trusting my abilities and seeing perspectives in my work, even when I got lost myself. He brought me into the Summer School on Computational Modelling of Cognition, and the International Max Planck Research School on the Life Course (LIFE), both of which had a large impact not only on the work in this thesis but my life in general.

Hauke has been my mentor for almost all of my years of study. Since I started as a HiWi in his lab during my Bachelor's studies, he welcomed me warmly in his group and always had an open ear for scientific and personal matters. The experiences in his lab have motivated me to pursue the SCAN master at Freie Universität and this doctorate. It was also Hauke who enabled and supported my visit to Scott Huettel's lab at Duke University, a valuable experience for me in many ways and the basis for one of this thesis' studies. I also thank Scott for giving me this exceptional opportunity and his contributions as a co-author of this study.

Much of this work is the product of intense collaboration with Armin Thomas, to whom I am also indebted. Thank you for all the mutual mentoring sessions, discussions, productive co-working and conference attendances from Berlin's coffee shops, and Freie Universität's basements to Bonn, Boston, and Toronto.

I would like to thank Ian Krajbich for providing the major foundations on which much of this thesis rests, insightful comments and discussions, and valuable contributions as a co-author.

I thank Angela Heine for familiarizing me with experimental work early on and sparking my ongoing interest in eye movements.

I thank all members of the Heekeren and Mohr labs I have met for many discussions and lots of feedback. I owe special thanks to Rasmus Bruckner not only for important feedback, on- and off-topic discussions, coffee breaks, and visits to the beer garden but also for letting me rely on his LaTeX template and practical know-how while finishing this thesis.

I am grateful to the IMPRS-LIFE program for giving me opportunities to view and present my work in a different light, to look (and travel) beyond my nose, and my fellow fellows for unforgettable experiences and friendship.

Finally, I want to thank my friends and family for their long-lasting support, trust, and patience. Mira, in particular, thank you for being my closest supporter and ally, over all the time, in highs and lows. You made everything better.

1 Summary

How do humans make simple preferential decisions, like decide what to have for breakfast at a hotel buffet? In contrast to the assumption of normative accounts of decision making, humans' preferences are often not stable but constructed at the time of choice and contingent on the decision makers' interaction with the environment. Recent evidence suggests that the allocation of visual attention during deliberation is closely linked to subsequent choices so that alternatives that are looked at longer are generally more likely to be chosen. Prior work has characterized the processes underlying simple decisions in terms of evidence accumulation over time, where momentary rates of accumulation depend on the decision maker's allocation of gaze, and a decision is made when accumulated evidence reaches a certain threshold. However, the generalisability of gaze-dependent accumulation remains unclear in multiple regards: It is not established how well gaze-dependent evidence accumulation describes individual decision makers' behaviour or to what extent the association between visual attention and choice varies between individuals. In addition, it is unclear to what extent the theory applies to behaviour in contexts where choices deviate from normative predictions more substantially. Finally, it remains debated whether visual attention causally influences or rather reflects the construction of preferences.

In this thesis, I address these questions across three empirical studies using computational models of the decision process. In Study 1 (Molter et al., 2019; Thomas et al., 2019), we first developed a novel gaze-dependent evidence accumulation model that allowed investigation of choice processes on the individual level. In addition, we published a corresponding Python toolbox to facilitate its use by others. Using this new tool, we demonstrated that gaze-dependent evidence accumulation accurately captures individuals' choice and response time data and associations with gaze allocation across four simple choice data sets. Our analysis revealed, however, that individuals strongly differed in the degree to which choices and gaze allocation were

linked and that this variability was associated with individual differences in choice consistency.

In Study 2 (Molter et al., 2021), we tested the gaze-dependent evidence accumulation framework in a multi-alternative, multi-attribute task involving choices between three risky gambles. The task was designed to elicit context effects in choice behaviour that challenge normative choice theories. These context effects describe preference changes depending on the set of available alternatives. We found not only choices but also decision makers' gaze allocation to be modulated by context, allowing a gaze-dependent evidence accumulation model derived from prior work to generalise to this more complex scenario.

In our preregistered Study 3 (Molter & Mohr, 2021b), we finally addressed the causal direction of the association between visual attention and choice. Participants made repeated choices between two risky gambles whose attributes were presented sequentially. This allowed the experimental control of the stimuli's presentation duration and order. Our results confirmed a causal influence of information search on preference construction. However, we identified presentation order, not duration, as the influencing factor, as alternatives presented last were chosen more frequently. Notably, causal order effects are only predicted by some gaze-dependent evidence accumulation models, highlighting potential for future theory development.

The studies generally confirmed positive associations between visual attention and choice and provided support for gaze-dependent evidence accumulation theories on the individual level and in more complex choice scenarios. However, our studies also revealed large individual differences and possible limitations of current computational models of decision making. We showed that accounting for those differences and implementing additional mechanisms like accumulation leak to predict acquisition order effects substantially improve prediction of individual choice behaviour.

Finally, I discuss these results on the active role of visual attention in the decision process and the theoretical model of gaze-dependent evidence accumulation in the broader context of constructed preferences and outline potential implications for the model-based analysis of choice and eye movement data.

2 Glossary

aDDM Attentional Drift Diffusion Model. An influential gaze-dependent evidence accumulation model (Krajbich et al., 2010; Krajbich et al., 2012; Krajbich & Rangel, 2011). It is an extension of the DDM which assumes that momentarily unattended alternatives' values are discounted by a constant factor (see gaze discount). 23–27, 29, 30, 32, 39–41, 45–47, 57–59, 62, 68, 69, 73, 74

attraction effect A context effect, where adding a third alternative that is similar but inferior to an existing one, results in a higher relative preference for the alternative that it is similar to. 6, 15, 19, 36, 53, 54, 62

BIC Bayesian Information Criterion. 42, 43

compromise effect A context effect, where adding a third alternative with extreme attributes results in a higher relative preference for the alternative that is then perceived as intermediate. 6, 15, 19, 36, 52

context effect A change in preference between two alternatives after a third alternative is introduced into the choice set. Typically studied using alternatives with multiple attribute dimensions. 5–7, 14, 15, 30, 33, 36, 39, 42, 52

DDM Drift Diffusion Model. An influential evidence accumulation model of simple, binary decision making (Ratcliff, 1978; Ratcliff et al., 2016). It assumes that relative evidence in favor of each of two response alternatives is accumulated in a noisy process until the relative evidence reaches a threshold corresponding to one of the alternatives. Notably, it predicts the identity and the time of the response first. 5, 23, 27, 58, 60

description invariance The normative tenet that preferences should not depend on the format in which choice alternatives are described. 14, 16, 17, 67

EUT Expected Utility Theory (Von Neumann & Morgenstern, 1947). 12, 15–18, 36, 40, 41, 53, 62, 71

evidence accumulation The principle of repeated, sequential sampling and integration of typically noisy evidence (in favour of choice alternatives). A key feature of most computational cognitive models of decision making including the Drift Diffusion Model, Decision Field Theory, the attentional Drift Diffusion Model, and the Gaze-weighted Linear Accumulator Model . 6, 18, 19, 23

gaze bias The empirically observed association of gaze duration and choice, such that alternatives that are looked at longer, are typically chosen more frequently. 22, 23, 31, 33, 38, 39, 58, 60, 61

gaze cascade A theory proposed by Shimojo et al. (2003). It posits that visual gaze and preference are associated in a positive reciprocal loop so that gaze towards an item increases preference for it, and preference for an item increases the likelihood of gaze towards it. The empirical effect where gaze towards the ultimately chosen item increases over the course of a decision is often called the gaze cascade *effect*. 22, 24, 28, 34, 55, 64

gaze discount The mechanism of a computational model of decision making by which value representations of momentarily unattended alternatives (or attributes) are discounted by a constant factor. 5, 23, 33, 40, 41, 43, 45, 57–62, 67, 74

GLAM Gaze-weighted Linear Accumulator Model. 40, 46–48, 51, 52, 54, 57–60, 62, 70, 73–75

MDFT Multialternative Decision Field Theory (Roe et al., 2001). An evidence accumulation model of multi-alternative, multi-attribute choice based on which can predict multiple context effects, namely attraction effects, compromise effects, and similarity effects. 19, 20, 30, 41, 53, 62, 71, 72, 74

perceptual choice The process of categorizing different stimuli of the environment, based on the subjective sensation of their physical features. 9, 19, 31, 35

preferential choice The process of deciding between multiple alternative actions based on idiosyncratic preferences. Also referred to as value-based decision making. 9, 10, 12, 19, 23, 31, 61

procedure invariance The normative tenet that preferences should not depend on the method by which they are elicited (e.g., the order of individually placed bids for risky prospects should match the preference order obtained from choice between those prospects). 14, 67

process model (of decision making). A theory of decision making that describes not only the decision outcome but the psychological (and possible neural) processes leading to a decision. 17, 18

PT Prospect Theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). 16–18, 36, 62, 68, 69, 71

RT response time. 16, 18, 23, 32, 39, 41, 45, 46, 49, 52, 66, 70, 74

similarity effect A context effect, where adding a third alternative that is similar to an existing one (and neither dominated nor dominating), results in a higher relative preference for the alternative that it is dissimilar to. 6, 15, 18, 19

WAIC Widely Applicable Information Criterion. 42, 48

3 Introduction

Our personal lives consist of series of decisions. Every morning, when we wake up, we decide what we want to have for breakfast, what we want to wear for the day, whether to take the car or the train to work, and so on. That is, after we decided when to set our alarm, whether to hit the snooze button (once? twice?), which groceries and clothes to buy, which career to pursue, and whether or not to buy the monthly subscription pass for the train. We make decisions all the time. It could be said, in fact, that every process between perceiving information from the world and acting is decision making, involving alternative courses of action at each point.

Not only are decisions everywhere in our lives, but the choices we make have a great part in what makes each individual different from – or similar to – other persons: Which newspaper to read, who to vote for, or whether to get vaccinated in times of a global pandemic or not.

The choices we make can broadly be divided into two classes: There is perceptual and preferential choice. Perceptual choices are choices about different states of our environment. For example, deciding whether the alarm is really ringing, the jar of cereal full or empty, whether the jacket is a very dark shade of blue or black. For these types of decisions, there exists an objective truth, and it's our task to identify it. Preferential choices, on the other hand, involve our own preferences: Whether to have the apple or the banana, whether the khaki or the green pants go better with the probably blue jacket, whether to take the train to work (even though it might be late and crowded), or the car (but there might be traffic, it's less environmentally friendly, and you might have to get gas anyway). In these situations, there is no right or wrong choice, but our personal and subjective preferences determine our actions.

The study of decision making is relevant not only because it is interesting to understand how humans perform this fundamental task, but also because the decisions we make can have important consequences: Whether to save for retirement or not, which investment to take, which home to buy? In addition, changes to decision making are at the core of many clinical conditions like addictions or eating disorders.

One aspect that has caught researchers' eyes more recently is that humans do not have omniscient insight into their environment but need to actively orient themselves towards different aspects of their environment, gathering information relevant to their decisions. For example, before one can decide whether to buy the daily, monthly, or single trip pass for the train, one has to glance at the menu of options on the vending machine screen and learn about those different options. Or, to choose a snack from a vending machine, the decision maker first has to find out which products are in stock. This information search often consists of *looking* and directing *attention* towards information in the environment relevant to our decision. There is also an intuitive indication that looking at different choice options and choosing them is somehow linked. Ask yourself, how often do you choose an item from a vending machine which you did not look at? How often did you choose one item from a vending machine while looking at another?

This thesis aims to address these kinds of questions and provide an increased understanding of the role of visual attention (i.e., where we look) during deliberation and what we end up choosing. It deals with simple preferential choice scenarios, but the hope is that findings might help address more profound decisions by increasing the understanding of the general mechanisms involved in decision making.

3.1 A framework of decision making

A basic framework to formalize different core components of the decision-making process is given by Rangel et al. (2008) and illustrated in the left half of Figure 3. In order to make value-based decisions, the decision maker is thought to perform multiple sequential steps. Note that this framework is a deliberate simplification to help structure investigation of the decision-making process. Both the sequential nature and the boundaries between the proposed operations are not rigid (Rangel et al., 2008).

First, the decision maker must form *representations* of the available alternatives, their attributes, and associated actions like button presses or arm movements that need to be performed to obtain them. Other relevant states of the environment that might be relevant to the decision (e.g., the amount of change in one's pocket, the other person waiting in line) have to be represented, too. Similarly, information internal to the decision maker, like hunger, thirst, or stress level, must be represented to be able to inform the decision. As many decisions involve visually presented choice alternatives (e.g., at the vending machine, or on the computer screen), those representations are likely to depend on decision makers' visual attention.

Second, many decision-making theories assume that choice alternatives and their associated actions need to be *valued*. The idea is that each action (e.g., choosing a candy bar from a vending machine) is assigned a scalar value, representing the expected benefits of taking it. Depending on the nature of the choice alternatives, different methods of valuation are assumed to be involved.

Third, when each action is assigned a value, the decision maker has to *select* which action to take, for example, by comparing the alternatives' values.

Fourth, after an action was made and an outcome is obtained, it is *evaluated*. If the decision maker chose the candy bar, the outcome of tasting and consuming the it is evaluated. In addition, actions can result in changes to the decision maker's internal or external states (like the chocolate affecting satiety), and these new states must also be evaluated.

Fifth, evaluations of the decision's outcomes can be used to *learn* and update all components of the decision mechanism: Has the set of available actions changed (is there another candy bar in the machine)? Have internal or external states changed (how much change is left)? Should actions be valued differently the next time (did the candy bar taste as good as expected)? Does the action selection mechanism require adjustment?

This thesis focuses on the mechanisms involved in *action selection* and *valuation* to some extent, and the role of visual attention, that is where decision makers look while making their decisions, in particular.

In the following sections, I will outline theoretical details and empirical findings referring to and expanding on the elements of this framework to motivate the open research questions this thesis aims to address.

3.2 Stable preferences

The study of preferential choice has a longer history in the field of economics than it does in psychology. Accordingly, economists had a significant head start in systematically defining concepts central to decision making like *value*, *preference*, and *risk*. Economic theories of individual decision making are, therefore, a natural starting point for psychologists' and neuroeconomists' (Glimcher & Fehr, 2014) investigation of choice, or the background against which their theories are contrasted.

While the field of economics has produced a myriad of choice theories, its most influential ones share a fundamental assumption: That decision makers have *stable* preferences which they seek to maximise (McFadden, 2001) and that preferences are "revealed" by their choices (Samuelson, 1938).

Another shared property of most economic models is their or *normative* character: Even though originally motivated by the desire to *describe* behaviour (Bernoulli, 1954, originally published in 1738) the neoclassical Expected Utility Theory (EUT) (Von Neumann & Morgenstern, 1947) is founded on a set of assumptions (or axioms) which decision makers *should* follow if they acted rationally.

Notably, regarding the *valuation* operation, the axiomatic approach of EUT provides a solution to the fundamental problem that the utility of goods and actions – denoting their desirability or subjective value – cannot be measured directly, but only inferred through choice: EUT shows that *if* decision makers' choice behaviour was compatible with its axioms, then their behaviour can be described *as if* they assigned a stable, numerical value or *utility* to each alternative and chose the one with the number. In case of uncertain outcomes, their utilities ought to be weighted by their probabilities of occurrence.

Action selection in standard utility theories is accordingly assumed to be as simple as identifying the highest-valued alternative. To account for the fact that human decision making is not deterministic (Luce, 1959; Rieskamp, 2008; Rieskamp et al., 2006), different probabilistic utility models have been developed: In random utility models (McFadden, 1973) *action selection* keeps its deterministic, utility-maximising status, but the underlying utilities are assumed to have a random error component leading to choice variability. Fixed utility models, on the other hand, assume that *action selection* itself is probabilistic (Luce, 1959) so that the highest-utility alternative is only most likely to be chosen, but not deterministically so.

Strictly speaking, economists do not insist that utilities are real and actually used by decision makers. They are "*as-if* models" (Friedman, 1953; Gigerenzer, 2020), agnostic about the real process underlying choice, and without relationship to psychological reality.

3.3 Constructed preferences

The *normative* and *as-if* status of economics' optimal choice models, together with a growing body of data incompatible with their fundamental axioms, prompted the search for more psychological and *descriptive* accounts of decision making.

Starting with Simon (1955) who introduced the concept of *bounded rationality*, many researchers adopted an "information-processing view" of decision making, focusing on the roles of perception, cognition, learning, and psychological representations of decision problems (Slovic, 1995). The general idea of these emerging theories is that human decision makers are biological systems, inherently limited in the resources they can operate on.

In contrast to the assumption of stable preferences which do not vary across occasions, and choices resulting from rational maximisation of utility, researchers considered preferences *constructed* at the time of choice, choices to be the product the decision makers' interaction with their environments, and thereby decision making a process highly contingent on task, context, and other factors (Lichtenstein & Slovic, 2006; Payne et al., 1992). Notably, theories of constructed preference explain a wide range of choice behaviour in conflict with the assumptions of rational choice models.

3.3.1 Preference reversals, frames, biases, and context effects

Rational models of choice make the basic assumption that preference orders should be identical ("invariant") across different ways of eliciting them (e.g., via choice or by ordering individual ratings). Similarly, different ways of presenting objectively identical choice alternatives should not result in changed preferences. Human decisions, however, violate both of these assumptions frequently and robustly: Lichtenstein and Slovic (1971) demonstrated that decision makers' preference order of two monetary lotteries reversed, depending on whether they chose between, or priced them individually, violating procedure invariance. The robustness of this violation was demonstrated to great effect by replicating it on a Las Vegas casino floor (Lichtenstein & Slovic, 1973). Violations of description invariance are illustrated in framing effects (Tversky & Kahneman, 1981), where decision makers' preferences reverse based on different descriptions of identical outcomes (e.g., saving a third of people from a deadly illness *vs.* letting the illness kill two-thirds of them). Many more so-called biases are documented, in which decisions deviate from rational utility models' predictions and are influenced by factors which those consider irrelevant.

Context effects Another set of decision-making patterns incompatible with most theories of stable preferences are context effects. These can be observed in choices with more than two alternatives that are described by multiple attributes. Consumer choices, for example, usually involve more than two alternative products, differing on multiple attributes like price and quality. Context effects refer to the change in relative preference between two alternatives after a third alternative is added to the choice set.

At least three context effects are in conflict with traditional models assuming stable alternative-wise utilities: The attraction effect describes an increased preference for an alternative after a similar but slightly inferior alternative is added (Huber et al., 1982). In the compromise effect, relative preference for an alternative is increased after the addition of a third alternative that makes it appear as intermediate (Simonson, 1989). The similarity effect (Tversky, 1972) predicts that adding an alternative that is similar to one of the original ones and similarly appealing will increase relative preference for the other, dissimilar alternative.

All context effects violate the rational principle of independence from irrelevant alternatives (Luce, 1959), which states that the relative preference between two alternatives should remain constant across different choice sets. In addition, the attraction effect violates the normative principle of regularity, which prescribes that preference for an alternative should only *decrease* after a choice set is enlarged.

Generally, models that presume independent valuation of each available alternative (e.g., fixed utility models) are unable to account for these observed violations (Rieskamp et al., 2006). For this reason, multi-alternative, multi-attribute choice scenarios provide rigorous testbeds for competing theories of preferential choice (Berkowitsch et al., 2014).

3.3.2 Theories of preference construction

Theories of constructed preference can be broadly classified by their *methodological* approach (Busemeyer et al., 2006): The first group of theories builds on classic utility-based models like EUT but includes modifications, for example, to the weightings of outcomes, to provide better descriptions of observed behaviour. The second stream of research, independent from traditional models, instead described choice behaviour in terms of heuristics, that is, simple rules that determine how information is processed and results in decisions. The third computational modelling approach, to which this thesis' methods are most closely related, describes preference construction in terms of dynamic and mathematically defined processes, which make precise predictions about choices and response times (RTs) in a given context (Busemeyer et al., 2006).

In addition, theories can be distinguished by their type of *explanations* for contingent decision making. Payne et al. (1992) identify two non-exclusive frameworks: In cost/benefit frameworks, contingent decision behaviour results from the consideration of different choice strategies' associated mental costs and benefits of application in a specific choice setting (Payne et al., 1988). In perceptual frameworks, on the other hand, contingent decision making is explained by more fundamental mechanisms associated with the representation of decision problems (Payne et al., 1992).

Prospect theory

In the domain of choices between alternatives with uncertain outcomes (risky choice), Prospect Theory (PT) (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) provides an accurate explanation for many observed deviations from EUT's predictions. This includes violations of description invariance as they are observed framing effects.

PT builds on the general framework of EUT, but includes several modifications to better describe observed behaviour: In a first "editing" phase, decision makers are assumed to form simplified *representations* of the decision problem at hand. In a second "evaluation" phase, not unlike in EUT, decision makers are assumed to assign *values* to each choice alternative, weighting their outcomes by their probabilities. In this phase, PT posits three major differences to EUT with respect to the *representation* of information. First, outcomes are thought to be represented with respect to a subjective reference point (e.g., the status quo or the decision maker's expectation). Second, objective outcome probabilities are assumed to be transformed into subjective decision weights in a non-linear fashion, where small probabilities are overweighted and large probabilities underweighted. Third, the value function which maps objective outcomes to subjective utilities is assumed to have a steeper slope and inverted curvature in the domain of relative losses compared to relative gains.

Like in EUT, however, *valuation* in PT involves alternative-wise weighting of outcome utilities by their subjective weights. Similarly, *action selection* follows the maximisation of subjective expected utilities over alternatives.

Notably, PT's changes in *representation* allow it to predict violations of description invariance and qualify it as a theory of constructed preference. As reference-dependence and non-linear probability-weighting are assumed to result from fundamental properties of humans' make-up, PT's explanations of contingent decision making are considered *perceptual* (Payne et al., 1992).

Adopting the general structure of rational utility models, especially with respect to *valuation* and *action selection*, PT, however, remains agnostic about the actual processes that decision makers use to make their choice. In this sense, PT stays close to the *as-if* status of rational utility models.

Heuristics

Heuristic approaches, in contrast, describe decision makers' information processing steps more directly. In specifying which information is used at which point in time (and characteristic of heuristics: which information is *not* used), they are models of the decision process.

Examples of heuristics include the priority heuristic (Brandstätter et al., 2006) which describes many deviations from EUT in risky choice by describing a list of "reasons" considered sequentially by decision makers. When a reason (e.g., the minimum gain of a lottery) is sufficiently diagnostic between alternatives, a decision is made, and no further reasons are considered. The elimination-by-aspects heuristic

(Tversky, 1972) similarly predicts which information is used by decision makers in decisions involving multiple alternatives with multiple attributes (e.g., choices between consumer goods). In Tversky's model, decision makers are assumed to sequentially consider different aspects (i.e., criteria relating to individual attributes; e.g., a maximum price) and rule out alternatives that exclude the selected aspect until only one remains and is chosen. Notably, this procedure is able to predict similarity effects.

It is noteworthy that heuristic models typically do not presume an intermediate and alternative-wise *valuation* step between representation and action selection. Instead of selecting the alternative with the highest *a priori* assigned value, *action selection* in heuristic models results from comparative and non-exhaustive processing of the alternatives' representations.

Contingent decision making emerges from heuristics in multiple ways: First, their comparative action selection makes them inherently context-dependent. Second, individual heuristics are often considered part of larger repertoires of decision strategies (Gigerenzer & Selten, 2001; Scheibehenne et al., 2013) which decision makers adaptively employ, depending on their costs and benefits in a given choice setting (Payne et al., 1992).

Dynamic computational process models of decision making

A third theoretical approach to preference construction is its description in terms of dynamic computational process models. (Busemeyer et al., 2006; E. J. Johnson & Ratcliff, 2014). In contrast to heuristic accounts, this approach involves the precise mathematical mapping from input variables like the choice alternatives' attribute values to the observable choice, and notably RTs. Unlike simple algebraic (E. J. Johnson & Ratcliff, 2014) mappings as in EUT and PT, however, these models also provide descriptions of the *process* leading to the decision. A central concept to this group of models is that of evidence accumulation, which describes the process of repeatedly sampling and integrating noisy evidence in favour of each choice alternative. In this framework, choices are made as soon as the accumulated evidence for one alternative reaches a certain threshold. Not only are many of these models inspired from biological systems (Roe et al., 2001; Usher & McClelland, 2001, 2004; Wang, 2002), but there is also a large body of research demonstrating evidence accumulation processes in neural data during perceptual and preferential choice, both in humans and nonhuman primates (Basten et al., 2010; Gold & Shadlen, 2007; Heekeren et al., 2008; Pisauro et al., 2017; Shadlen & Newsome, 2001)

An example of a dynamic computational process model of decision making is Multialternative Decision Field Theory (MDFT) (Roe et al., 2001). Like heuristic approaches, this evidence accumulation model does not assume alternative-wise *valuation*. Instead, it posits the computation of choice alternatives' *valences*, derived from attribute-wise comparisons with other alternatives. The attribute selected for this comparison is stochastically determined by a so-called "attention" mechanism (see below).

The model further assumes leaky evidence accumulation, where the influence of early acquired evidence diminishes until evidence for one alternative reaches a threshold and a choice is elicited. In addition, accumulated evidence for different alternatives are assumed to *inhibit* each other, depending on the similarity of the alternatives' attributes.

MDFT makes quantitative predictions about similarity, compromise and attraction effects and, illustrating the additional value of computational process models, also how they change under different levels of time pressure (Roe et al., 2001).

Notably, these models' explanations can be described as *perceptual* assume contingent decision behaviour to result from the dynamic interplay of elemental processes within the decision-maker.

3.4 Attention

Attention broadly describes the set of mental processes that serve functions of selection, modulation, and maintenance of information in the face of limited cognitive resources (Chun et al., 2010; Krauzlis et al., 2021). With researchers focusing on the *psychological* factors influencing decision making and the general limitations to humans' information processing capacity, the concept of *attention* took a more prominent role in theories of constructed preferences (Weber & Johnson, 2009). Unlike rational models, which trivially assume that decision makers should consider all relevant and disregard all irrelevant information, attention therefore serves to provide certain information in a given choice context with additional weight, and changes the resulting decisions accordingly (Russo, 2019). The specific mechanisms attributed to attention, however, vary considerably:

Many computational *process models* of multi-alternative, multi-attribute choice contain mechanisms labelled attention (Busemeyer et al., 2019): In MDFT and other models (Usher & McClelland, 2004), the mechanism that stochastically switches between momentarily evaluated attributes is called attention. In the associative accumulation model (Bhatia, 2013), the weights given to different attribute dimensions are determined by a mechanism called attention. Loss attention (Yechiam & Hochman, 2013) describes a general increase in alertness in the face of possible losses, highlighting yet another facet of the attention construct (Petersen & Posner, 2012).

This overly broad use of attention as an explanatory device has, however, drawn criticism for the many different functions and meanings associated with the term (Chun et al., 2010; Hommel et al., 2019; Krauzlis et al., 2021), the risk of circular arguments (Hommel et al., 2019), and the lack of direct measurements (Krauzlis et al., 2021).

3.4.1 *Visual* attention in decision making

In contrast, a more concrete and measurable approach to investigating decision makers' behaviour considering limited processing capacity is to focus on *overt visual attention*. Overt visual attention describes the mechanism by which an agent selects one location for preferred processing over others by moving their eyes towards this location.

Generally, eye movement recordings can be used to measure overt visual attention (Rayner, 2009). Even though attention can also be deployed covertly (i.e., without moving the eyes; Orquin & Holmqvist, 2018) and short periods of dissociation between visual fixations and attention precede the shift of gaze (Liversedge et al., 2011), these misalignments are unlikely to be strategic when decision makers can inspect information freely (Rayner, 2009).

Intuitively, it is plausible to assume that visual attention has at least some relevance in the decision making process: Unlike many theories implicitly assume, decision makers typically do not have the ability to instantaneously form a full *representation* of all relevant information. Standing in front of a vending machine or a supermarket shelf, for example, decision makers first have to identify which products are in stock. This is often done by *sequential* and *visual* inspection. Even when all alternatives are identified, casual observation shows that decision makers do not choose with their eyes closed, but when deciding between candidate alternatives, they shift their gaze between them before making the decision.

Empirical findings Many empirical studies have investigated, mostly using eye tracking, the factors that influence the way decision makers allocate their visual attention (for a review see Orquin & Mueller Loose, 2013). These studies identify both bottom-up, stimulus-driven, and top-down, goal-directed factors (Corbetta & Shulman, 2002; Orquin & Mueller Loose, 2013). Notably, some drivers of visual attention are also associated with changes in choice behaviour.

First, the position of choice items determines whether and when they will be attended. Generally, items positioned at the top and left of a display are attended earlier and more (Chandon et al., 2009; Krajbich et al., 2010; Krajbich & Rangel, 2011; Thomas et al., 2021), which has been associated with decision makers' learned reading direction. In addition, centrally displayed items are attended more (Chandon et al., 2009) and earlier (Thomas et al., 2021) and also chosen more frequently (Chandon et al., 2009).

Next, more visually *salient* items, referring to their distinctiveness against the background, and subsuming features like contrast and colour (Itti & Koch, 2000, 2001), are attended and chosen more (Milosavljevic et al., 2012; Towal et al., 2013). Similar effects have been reported for the physical size of choice options (Chandon et al., 2009; Thomas et al., 2021).

In tasks where decision makers provide estimates of single alternatives' value (e.g., by rating them individually), items of higher value are attended more (Gluth et al., 2020; Krajbich & Rangel, 2011; Thomas et al., 2021; Towal et al., 2013), and increasingly so over the course of deliberation (Thomas et al., 2021).

Relatedly, items that are ultimately chosen are attended more in total (S. Fiedler & Glöckner, 2012; Folke et al., 2016; Glöckner & Herbold, 2011; Isham & Geng, 2013; Kim et al., 2012; Krajbich et al., 2010; Krajbich & Rangel, 2011; Stewart et al., 2016; Tavares et al., 2017), a finding referred to as gaze bias. The association between gaze direction and choice also increases over the period of the choice (S. Fiedler & Glöckner, 2012; Shimojo et al., 2003). This gaze-cascade effect (or late onset bias; Mullett & Stewart, 2016) has been theorized to result from a positive feedback loop between mere exposure effects (i.e., looking at an alternative increasing preference for it; Zajonc, 1968) and preferential looking (B. A. Anderson et al., 2011), but can also emerge without such a loop (see below; Mullett & Stewart, 2016).

Finally, decision makers' last fixation before making a choice is mostly directed towards the chosen alternative (Krajbich et al., 2010; Krajbich & Rangel, 2011).

In summary, a wealth of empirical data has established a generally positive association between overt visual attention and choice. Attention towards an alternative is associated with a higher likelihood of choosing it and an increase of this association's strength over the time of choice.

3.4.2 Gaze-dependent evidence accumulation

Findings of gaze biases and casual observations that decision making involves repeated shifts of gaze between choice alternatives motivated the development of computational models of *action selection* that explicitly integrate decision makers' patterns of visual attention into the choice process. These gaze-dependent evidence accumulation models formalize the empirically observed association between gaze and choice. They are based on the DDM (Ratcliff, 1978; Ratcliff et al., 2016), a highly influential evidence accumulation model of decision making. In the DDM the relative evidence in favour of each of two possible responses is accumulated until it reaches one of two decision thresholds, which triggers the corresponding response. While the model was initially developed to explain memory retrieval (Ratcliff et al., 2016) it was also shown to predict choice and RT data in preferential choice by defining the average rate of accumulation (the drift rate) as a difference of item values (Milosavljevic et al., 2012).

In contrast to the DDM, gaze-dependent evidence accumulation models make the additional assumption that the momentary drift rate of evidence accumulation depends on which alternative is currently fixated by the decision maker (Figure 1).

In the influential attentional Drift Diffusion Model (aDDM) (Krajbich et al., 2010; Krajbich & Rangel, 2011) evidence accumulation for an alternative is assumed to be discounted by a constant factor while another item is fixated (Figure 1). This gaze discount mechanism lets the model explain the observed gaze bias and provide precise predictions of choices RTs and eye movement data (Ashby et al., 2016; Cavanagh et al., 2014; G. Fisher, 2017; Gluth et al., 2020; Gluth et al., 2018;

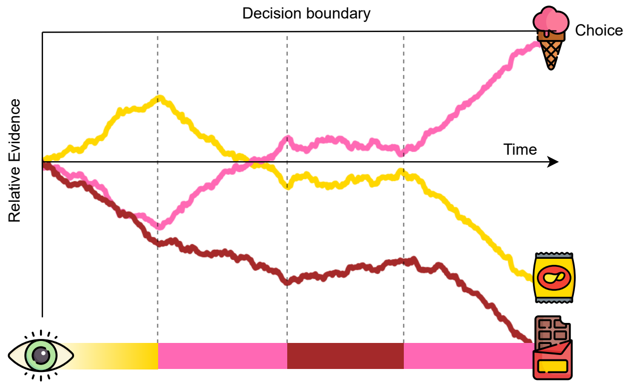


Figure 1. Gaze-dependent evidence accumulation In the multialternative attentional Drift Diffusion Model (Krajbich & Rangel, 2011) decision makers accumulate relative evidence in favour of each available alternative (ice cream, crisps, and chocolate). Crucially, momentary accumulation rates depend on the allocation of the decision maker's gaze so that evidence in favour of an item accumulates faster while it is fixated. A choice is made, once the relative evidence in favour of one alternative reaches a given threshold, the decision boundary. In this example, the decision maker chose to eat ice cream.

Krajbich et al., 2010; Krajbich et al., 2012; Krajbich & Rangel, 2011; Towal et al., 2013).

Specifically, the model correctly predicts an increased likelihood of choice for alternatives that are looked at longer due to evidence for non-fixated alternatives being discounted. Notably, it also correctly predicts that this effect should be reversed for choices between aversive items (Armel et al., 2008). It also accounts for the fact that decision makers typically choose the alternative they fixate last, as fixated alternatives are more likely to cross the decision threshold than unfixated, discounted alternatives (unless their value is much lower). Last fixations are also correctly predicted to be shorter, because they get interrupted when the decision threshold is reached. Similarly, the aDDM and other gaze-dependent evidence accumulation models with a relative choice criterion provide an account for the gaze-cascade effect, even without assuming preference-driven gaze (Mullett & Stewart, 2016).

Multiple studies found neurobiological correlates of the aDDM's assumed discounted value signals: Using electrophysiology Hunt et al. (2018) and McGinty et al. (2016) recorded fixation-dependent value signals in monkeys' orbitofrontal cortex, while the animals freely viewed reward-associated cues. In humans, Lim et al. (2011) found similar fixation-dependent relative value signals during two-alternative choice with enforced fixation patterns in the ventromedial prefrontal cortex and the ventral striatum using functional magnetic resonance imaging (fMRI).

Application of the aDDM to data from simple two- and three-alternative snack food choices revealed that decision makers, on average, appear to discount the value of unattended alternatives roughly by a factor of one-third (Krajbich et al., 2010; Krajbich & Rangel, 2011).

Notably, these findings were obtained from applying the aDDM to *group-level* data. As the model aims to describe and explain *individual* decision makers' behaviour, however, it is essential to also demonstrate its ability to explain data on this *individual* level. See Box 1 for an example of how inferences about individuals' behaviour from group level data can be misleading.

Box 1| Inferences from aggregate data

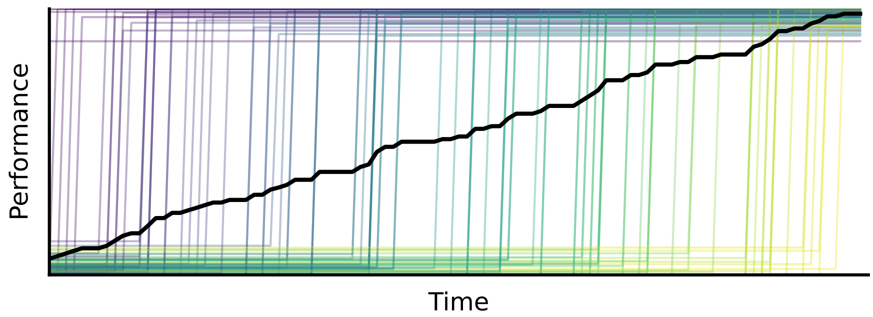


Figure 2. Fictional data of people solving a puzzle. Each colored line marks a single person's ability to solve a puzzle over time. The black line shows the group-average performance evolving over time. Example adapted from Hayes (1953).

Suppose a friend hands you a puzzle made up of wooden blocks and rings connected by a string, seemingly impossible to disentangle. She asks you to free one particular wooden ring from the string. After fidgeting for three minutes, you have an epiphany and see the one required move and solve the puzzle. Knowing the solution, you could do the puzzle again in an instance.

Now imagine many people solving the puzzle. Most people will find the solution eventually, and for most people, the puzzle will be easy after they solved it the first time. It is, however, likely that people differ in how fast they find the solution.

Fictional data from this thought experiment are plotted in Figure 2, where each coloured line shows the ability of a single person to solve the puzzle. Note how they all start at the bottom but make a step towards the top, when a person had their epiphany and found the solution. The black line shows the average performance across puzzlers at each point in time. Notice how the average appears to increase constantly and smoothly over time (because the time at which people's performance steps up varies), while this is not true of any individual performance curve.

A theoretical problem ensues when the aggregated group data is used to make inferences about the constituting individuals' behaviour: A theory positing smooth and incremental learning describes the aggregated data well but fails to describe any individual. In general, inferences from aggregate to individual data are only valid under specific circumstances (A. J. Fisher et al., 2018; Molenaar, 2004). Therefore it is important to test theories of individual behaviour on data matching this level.

Given the significant challenge which context effects pose to many models of decision making, it also remains unclear, however, whether the aDDM will also account for choices in scenarios involving more than two alternatives with multiple attributes, where these can emerge. One possibility in which gaze-dependent models could predict context-dependent choices would be if the allocation of gaze itself would change depending on the set of available alternatives.

Practical limitations Despite its predictive accuracy and intuitive appeal, empirical application of the aDDM, especially on the individual level, is held back for multiple reasons: First, there exists no off-the-shelf solution like those available for the application of the DDM (e.g., Vandekerckhove & Tuerlinckx, 2008; Voss & Voss, 2007; Wagenmakers et al., 2007; Wiecki et al., 2013). Second, due to its fixation-dependence, there is no analytical solution to its first passage density distribution (the predicted distribution of response times, conditioned on choice). Therefore custom implementations of the model need to rely on processing-intensive simulations. Third, as the accumulation process is assumed to depend on eye movements, these simulations have to include the simulation of decision makers' fixation paths. This was solved for simple two- and three-alternative choice (Krajbich et al., 2010; Krajbich & Rangel, 2011), but becomes exponentially more difficult for more complex (e.g., multi-alternative, multi-attribute) choice scenarios.

3.4.3 Causal direction

Another more profound question surrounding the aDDM is that of causality. Even though gaze-dependent computational models of choice are usually interpreted this way, they technically do not imply any causal direction of the association between gaze duration and choice but merely formalize their association. It would, for example, be possible that a third unknown variable causes both shifts in gaze and modulations of the evidence accumulation process. In this case, changes in gaze would not necessarily affect the choice process, as the unknown third variable could have remained unchanged.

The question of whether aspects of information search, like the duration for which choice alternatives are inspected, *causally* affect choice behaviour, however, has important implications: If they did, choices could be subject to irrelevant, external factors influencing information search. On the one hand, this would provide opportunities to externally influence decisions using choice architecture (Thaler & Sunstein, 2009) that considers human information search more specifically. On the other hand, identifying visual attention as a causal and constructive element in decision making would emphasize a theoretical perspective to understanding decision biases and invariances that can build on a large body of prior work concerned with the allocation of visual attention.

Multiple studies addressed this question of causality by experimentally controlling aspects of information search. Some studies presented alternatives sequentially, thereby controlling the total duration for which alternatives were presented (Armel et al., 2008; Shimojo et al., 2003). Others used gaze-contingent decision prompts (Liu, Lyu, et al., 2020; Pärnamets et al., 2015; Sui et al., 2020; Tavares et al., 2017), where participants can freely inspect alternatives but are prompted to make a choice as soon as their viewing patterns (recorded using eye tracking equipment) fulfil preset criteria (e.g., the decision maker looked at one alternative for 500 ms longer than the other). All studies confirm causal effects of viewing or presentation duration on choice, in line with causal interpretations of gaze-dependent evidence accumulation models (but see Newell & Le Pelley, 2018).

Notably, these studies focused on causal effects of alternative-wise viewing *durations* on choice. Other aspects of information search like the *order* in which information is acquired, however, are also associated with choice, as indicated by last-fixation and gaze-cascade effects. In addition, visual attention towards *attribute* dimensions (e.g., outcomes and probabilities in risky choice) have also been associated with differences in choice (e.g., S. Fiedler & Glöckner, 2012; Glöckner & Herbold, 2011; Kim et al., 2012), but are investigated less frequently.

Initial studies found choices to be causally influenced by the location of the last fixation (Liu, Zhou, et al., 2020) and by viewing duration of individual attributes (Liu, Lyu, et al., 2020; Sui et al., 2020).

Furthermore, associations between alternative-wise viewing duration and choice might have been confounded with viewing *order* in prior work: Chosen alternatives are often looked at longer *and* last, especially in fast-paced decisions which are often made with one, two, or three fixations (e.g., Krajbich et al., 2010). Notably, this potential problem can also arise in experimental designs using gaze-contingent decision prompts – depending on the conditions triggering the choice prompt.

In sum, there is growing evidence of a directed causal effect of alternative-wise gaze-duration on choice, but causal effects of other aspects of information search are only recently being revealed, and their independent contributions remain challenging to distinguish.

3.5 Summary

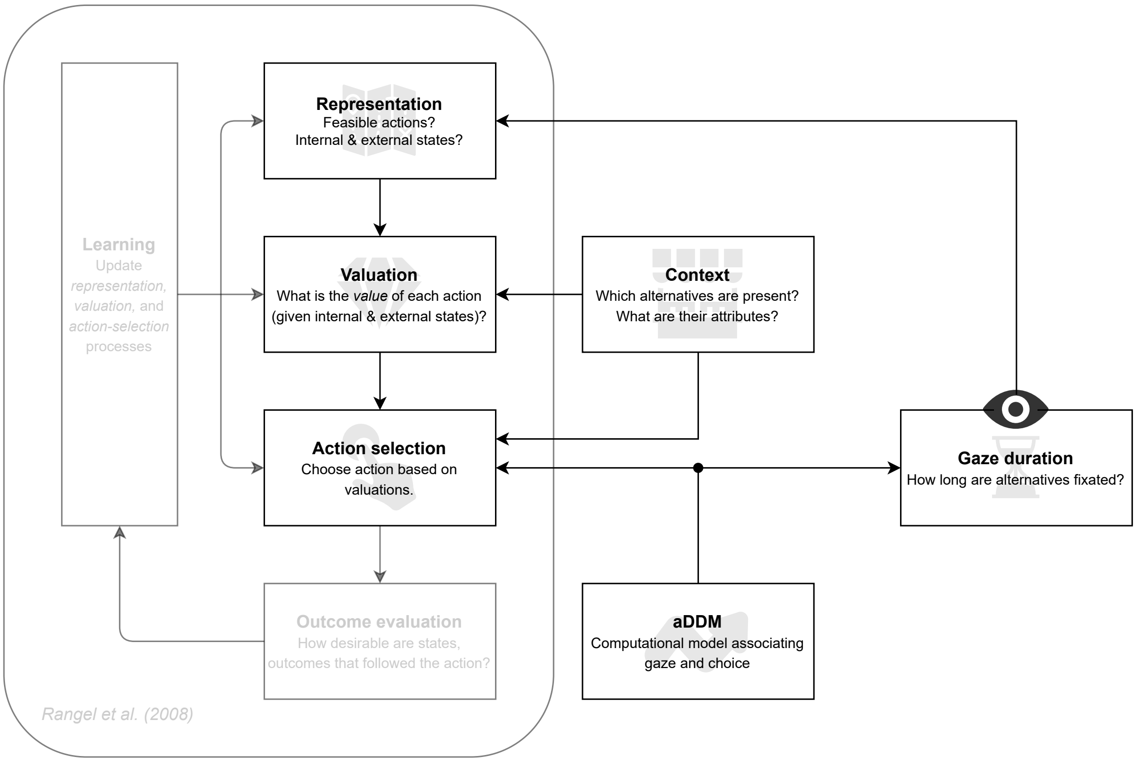


Figure 3. An expanded framework of decision making. Decision makers first have to form *representations* of internal and external states, including available alternatives and their attributes, and associated courses of action. Next, many theories of decision making assume a *valuation* step where each alternative is assigned a scalar value that denotes its expected benefits. The decision maker then *selects an action*, *evaluates its outcomes* and uses new information to update all processes during *learning*. Context effects illustrate that *action selection* depends on the context of available alternatives, with some models assuming context-dependent *valuation*. Finally, visual attention is linked to *action selection*, with the aDDM providing a theoretical account of this association. Adapted from Rangel et al. (2008).

At this point, the initially presented framework has been expanded in two main aspects (Figure 3): First, from the literature on the constructive nature of preferences, the influence of the *context* given by the set of available alternatives is now included more explicitly. These context effects demonstrate clearly that the set of available alternatives can affect action selection, with MDFT and related computational theories providing mechanistic explanations for this effect. Note that an edge is also drawn from context to valuation, as some models of context-dependent choice (especially normalization accounts like Louie et al., 2013; Soltani et al., 2012) locate the influence of context more on this level, rather than action selection. Second, the observed link between action selection and *visual attention*, especially the duration for which alternatives' are viewed, has been included. Gaze-dependent evidence accumulation in form of the aDDM provides a theoretical framework explaining this connection (on the group level, as noted earlier). In addition, visual attention is also assumed to affect decision-relevant representations.

4 Research questions

The goal of this dissertation is to provide an improved understanding of the role that visual attention plays in individuals' choice behaviour. Specifically, I investigate (i) interindividual variability in the strength of the association between gaze and choice in simple preferential and perceptual choice scenarios, (ii) the predictive performance and functional form of gaze-dependent evidence accumulation in context-dependent choice, and (iii) the causal effect of different aspects information presentation on choice processes.

Addressing individual variability in the association between gaze and choice, I aim to answer the questions:

1. *How can individual gaze-bias mechanisms be investigated efficiently using computational modelling?*

2. *Does gaze-dependent evidence accumulation capture simple choice behaviour on the individual level?*

3. *Which interindividual differences exist in gaze-bias strength and how do they relate to individual differences in choice behaviour?*

With respect to generalization of gaze-dependent evidence accumulation to context-dependent choice, I address the question:

4. *To what extent can gaze-dependent evidence accumulation explain complex multi-attribute, multi-alternative risky choice?*

Lastly, regarding the causal effects of presentation characteristics on choice, I investigate:

5. *Can binary risky choices be causally influenced by external control of presentation duration?*

6. *Can binary risky choices be causally influenced by external control of presentation order?*

Rationales and hypotheses

Question 1. *How can individual gaze-bias mechanisms be investigated efficiently using computational modelling?* Application of the standard model of gaze-dependent accumulation (the aDDM) to individual-level data posed a practical and technical challenge for multiple reasons, such as the lack of an analytical solution to its first-passage time distribution, and the need to simulate fixation patterns. Existing approaches to work around these practical difficulties (e.g., Cavanagh et al., 2014; Smith et al., 2019) did not apply to multi-alternative choice. In Thomas et al. (2019) and Molter et al. (2019), we therefore set out to develop a gaze-dependent accumulation framework that is computationally tractable and applies to choice settings with an arbitrary number of alternatives.

Question 2. *Does gaze-dependent evidence accumulation capture simple choice behaviour on the individual level?* Prior work (Krajbich et al., 2010; Krajbich et al., 2012; Krajbich & Rangel, 2011) could demonstrate that gaze-dependent accumulation provides a parsimonious account of choice, RT and gaze data and their interactions. However, these studies focused on the group level, leaving open whether gaze-dependent accumulation also captures *individual* participants' behaviour, since a description of aggregate group data does not necessarily characterize data of its constituents well (see Box 1). To address this question, in Thomas et al. (2019) we tested a novel gaze-dependent accumulation model on data from individuals across four existing data sets, and performed systematic model comparisons with competing gaze-*in*dependent theories on the individual level. We expected gaze-dependent evidence accumulation to describe individuals' choice behaviour better than competing models without gaze-dependence.

Question 3. *Which interindividual differences exist in gaze-bias strength and how do they relate to individual differences in choice behaviour?* Similarly, while prior work (Krajbich et al., 2010; Krajbich et al., 2012; Krajbich & Rangel, 2011) has focused on the average strength of the gaze discount across participants, variability therein between participants was not addressed. As a result it remained unclear whether individuals were well described by the group average. In addition, individual estimates of the gaze discount are necessary to investigate individual differences in gaze bias strength, and establish associations with other sources of interindividual variability. We addressed these issues in Thomas et al. (2019), by estimating individual gaze discount factors and investigating associations with other metrics of simple choice behaviour. Based on the limited evidence from prior work (Krajbich et al., 2010), we expected to find substantial variability between individuals, such that some individuals' choices were strongly associated with their gaze, whereas others' would not.

Question 4. *To what extent can gaze-dependent evidence accumulation explain complex multi-attribute, multi-alternative risky choice?* Having confirmed that gaze-dependent accumulation is a good model of individuals' choice behaviour in simple two- and three-alternative choice in Thomas et al. (2019), we next aimed to test, whether this finding holds when choices are made in more complex scenarios, that have long proven to be a challenge for traditional models of decision making. In Molter et al. (2021), we therefore tested different forms of gaze-dependent evidence accumulation models in a task where participants' choice behaviour was subject to context effects. We hypothesized that gaze-dependent accumulation could provide a process-oriented model of context-dependent choice, even in the presence of context effects, due to the context-dependent allocation of gaze.

Question 5. *Can binary risky choices be causally influenced by external control of presentation duration?* Based on prior experimental work (e.g., Armel et al., 2008; Shimojo et al., 2003), and predictions from causal interpretations of gaze-dependent accumulation models including the one used in Molter et al. (2021) (Glickman et al., 2019; Krajbich et al., 2010; Krajbich & Rangel, 2011), in Molter and Mohr

(2021), we experimentally tested the causal influence of different aspects of stimulus presentation on choice. We hypothesized that presentation duration affects choice, such that longer shown alternatives, and alternatives with better values on longer shown attributes would be chosen more frequently.

Question 6. *Can binary risky choices be causally influenced by external control of presentation order?* In addition, based on empirical associations between choice and within-trial acquisition order as in gaze cascade and last-fixation effects, we hypothesized that presentation order similarly affects choice, such that last presented alternatives and those which have better values on last shown attributes would be chosen more frequently.

5 General methodology

This section provides an overview of the general methodology used in the empirical and model development studies that are part of this thesis. Specifically, I will outline the general structure and characteristics of the different decision-making tasks, the eye tracking procedures, and the general principles that guided computational modelling of decision making. Readers are referred to the original publications for comprehensive descriptions of the methods used.

5.1 Behavioural tasks

All studies that make up the body of this thesis investigate human decision making behaviour measured in experimental tasks, where participants repeatedly indicate their preference between different choice alternatives shown on a computer screen.

In Thomas et al. (2021), we analysed four existing data sets from prior work using different behavioural tasks: The data from Krajbich et al. (2010), which included 39 participants making 100 choices between pairs of 70 different snack foods, the data from Krajbich and Rangel (2011), which included 30 participants each making 100 choices between three snack food items, and the data from Folke et al. (2016, Experiment 2), where 24 participants each made 144 choices between three snack food items. Lastly, with the data from Tavares et al. (2017, Experiment 1), Study 1 also included data from a different decision making domain, namely perceptual choice. Here, 25 participants performed 1344 trials across four sessions, in which they decided which of two slanted line segments closer matched a target exemplar.

Snack food experiments included additional valuation tasks, where participants indicated liking ratings (Krajbich et al., 2010; Krajbich & Rangel, 2011) or their willingness-to-pay (Folke et al., 2016) for each item individually. These value estimates served to assess the overall quality of choices (whether highly valued items are chosen over others) and are important inputs to the computational models used to describe choice behaviour.

In Molter et al. (2021) and Molter and Mohr (2021), we collected new data from participants choosing between multiple risky prospects, each described by a probability p to win an amount m and nothing otherwise. The use of risky prospects as choice alternatives has a long history in the literature on judgment and decision making (e.g., Allais, 1953; Tversky, 1972). Unlike snack food items used in Study 1, which are represented by an image of the item on the screen, risky prospects have two explicitly communicated attributes (probability and payoff) and therefore act as multi-attribute stimuli. Additionally, they provide a high degree of control over their attribute values (in contrast, for example, to snack foods' attributes like taste, mouthfeel, etc.) and a straightforward way to incentivise choices, even in an online setting. Finally, they serve as prototypical stimuli to the most influential theories of value-based decision making, namely EUT and PT. Molter et al. (2021) included 40 participants, collected in the laboratory at Freie Universität Berlin. In Molter and Mohr (2021), we aimed to collect a larger sample. Therefore this study was conducted online. This data set includes 179 participants.

Risky prospects in Molter et al. (2021) were tailored to individuals' risk preferences using an integrated adaptive staircase procedure. Importantly they were also designed to elicit context effects by adding different third prospects to a constant core set of two alternatives: Asymmetrically dominated prospects (that had lower p and m than one other available prospect) were added to elicit attraction effects. Extreme prospects (with very high p and low m or vice versa) were added to make one of the original options appear intermediate, and thereby elicit compromise effects.

While choices in Thomas et al. (2019) and Molter et al. (2021) were made without time limit and participants could freely inspect all choice alternatives, Molter and Mohr (2021) used a different approach to investigate the causal direction of the association between these variables and choice: Here, each trial was divided into a presentation- and a choice phase. Participants first learned about the two prospects' attributes in a sequential presentation over five seconds and were then prompted to indicate their choice within three seconds. Crucially, presentation duration, order, and format were experimentally controlled and varied, such that individual alternatives or attributes were shown longer and/or last in the sequence.

All value-based choices tasks (i.e., excluding the perceptual-choice data set in Thomas et al., 2019) involved repeated incentivized choices that had real consequences to the participants. This ensured that participants were motivated to choose according to their preferences. In snack food choices, they were asked to refrain from eating for three (Krajbich et al., 2010; Krajbich & Rangel, 2011) to four (Folke et al., 2016) hours prior to the task and could only eat an item they chose in a randomly selected trial afterwards. In risky choice tasks, one risky prospect the participant chose in a randomly selected trial would be played out for a real money bonus, paid in addition to the base compensation. This procedure was used to prevent participants from building "portfolios" of prospects with their choices (i.e., strategically combining prospects with different winning probabilities), yet ensuring that they had an incentive to treat each trial as if it was relevant for their bonus payment.

In all tasks of Thomas et al. (2019) and Molter et al. (2021), participants' eye movements during the decision phase were recorded (see Eye tracking). Importantly, all available choice alternatives (and their attributes in Molter et al., 2021) were presented on different positions on the screen, and there were no implicit default alternatives that were not represented on the screen. This allowed eye movement recordings to be linked to information pertaining to different choice alternatives (and their attributes).

In summary, the different studies measured choice behaviour across multiple decision-making domains, multiple set sizes, and different stimulus categories with both implicit and explicit representations of multiple attributes (including designs to elicit complex context-dependent choice), allowing for the analysis of gaze-bias effects across a wide range of different choice scenarios. Additionally, one task was specifically designed to investigate causal effects by experimentally controlling multiple aspects of stimulus presentation.

5.2 Eye tracking

Eye tracking, that is, the recording of eye movements, was used in multiple studies of this thesis to obtain process data about the duration for and order in which participants attended to choice options and their attribute values during the course of decision making. Experiments from Thomas et al. (2019) and Molter et al. (2021) collected eye movement data using different video-based eye tracking systems (Holmqvist, 2011), with sampling rates ranging from 50 Hz to 1000 Hz.

All eye tracking analyses focused on eye fixations, referring to the periods in which participants' gaze rests on a screen location and visual information is brought into the eyes' fovea before gaze is shifted to the next location (Liversedge et al., 2011). Saccades (the fast, jerky eye movements between fixations), blinks, and other types of eye movements were discarded. Fixations were detected from the raw stream of timestamped gaze coordinates using specific algorithms (Holmqvist, 2011) and assigned to the different choice options (and their attributes in Molter et al., 2021) on the screen by matching their positions to areas of interest constructed around the stimulus positions on the screen. Finally, multiple different variables (e.g., the overall duration a choice alternative was fixated in a trial) pertaining to the duration and sequence of information search during deliberation were constructed and analysed.

In general, the use of eye movements for analyses of choice behaviour served two purposes: First, we used eye movement data to describe and characterise participants' information search behaviour during deliberation. Since we were particularly interested in gaze bias effects, these analyses also included descriptions of the association between gaze and choice. For example, in Thomas et al. (2019), we devised a behavioural measure of gaze influence, quantifying the change in choice probability associated with looking at an item longer than others. This measure was then used to quantify individual differences in gaze bias strengths. In Molter et al. (2021), we computed relative dwell times (i.e., the relative amount of total fixation time) towards each choice alternative to test whether — like their choices — participants' gaze was subject to context effects. In addition to regressing dwell times onto trial characteristics to gain a more complete understanding of their distributions, we analysed the *direction* of information search. Similarly, other descriptive measures were used to evaluate *absolute* model performance in Thomas et al. (2019) and Molter et al. (2021) (see Model comparison).

Crucially, for most model-based analyses of this thesis, eye tracking data also served as additional input to the computational models. Gaze-dependent accumulation models (including the aDDM and related models) posit that decision variables depend on momentarily fixated alternatives or attributes. The application of these models, therefore, involved empirically measured eye tracking data. Specifically, the model developed in Thomas et al. (2019) and Molter et al. (2019) uses relative gaze durations towards each alternative to discount and weight value signals. Similarly, models in Molter et al. (2021) used the sequence of fixated alternatives and attributes in each trial to predict choices.

5.3 Computational modelling of decision making

Across the studies of this thesis, computational modelling of decision making was a main tool of analysis. Computational models of decision making are quantitatively precise theories about how decisions are made. In contrast to verbally stated theories, they involve mathematical definitions of all relevant external and internal information (e.g., stimulus attributes like outcomes and their probabilities, or idiosyncratic values of choice items), their representations, and computational algorithms acting on them to generate observable choice (and often RT) behaviour.

In the context of this thesis, the use of computational models serves multiple purposes: First, studies involved the comparison of different competing computational models with the goal to identify the most likely process underlying individual decision behaviour in the different experiments. Second, computational modelling was used to quantify individual differences in the expression of particular mechanisms of the decision making process: In Thomas et al. (2019) and Molter et al. (2021), we estimated gaze discount parameters, quantifying the degree to which unattended information is discounted in the accumulation process. Third, acting as theories of decision making, we derived hypotheses about choice behaviour from the class of gaze-dependent accumulation models: In Molter and Mohr (2021), we tested model-predicted effects of presentation duration and order on risky choice behaviour.

In the next sections, I provide additional details about the individual steps of the computational modelling analyses.

5.3.1 Specification of models and model spaces

The first step for a model-based analysis of decision making is to define the set of models to be included (the *model space*). The selection of models is consequential as it defines and limits the possible results from subsequent model comparison analyses. Depending on the specific goals, different approaches to this were taken in the studies of this thesis:

In Thomas et al. (2019) and Molter et al. (2019), we developed a novel gaze-dependent accumulation model (the Gaze-weighted Linear Accumulator Model (GLAM)) that builds on the aDDM from prior work (Krajbich & Rangel, 2011), but is more suited for the application to data from individual participants. The model space included this novel model and a restricted variant without gaze-dependence, allowing us to test the presence of gaze discount mechanisms on an individual level.

In Molter et al. (2021), we first aimed to compare performance of gaze-dependent accumulation in context-dependent risky choice against a set of established reference models of risky (EUT; Von Neumann & Morgenstern, 1947) and context-dependent choice (MDFT; Roe et al., 2001), and multiple control models. In a second step, we tested the *a priori* defined gaze-dependent accumulation model against a systematically defined, large space of variants: Here, different model *mechanisms* and their different implementations were identified (e.g., from prior work) and exhaustively combined to create novel hybrid models that mix and match individual mechanisms. This approach can generate a much more dense and complete space of models to select from than a selection of *a priori* defined models and has the added advantage that the relative contribution to overall model fit of individual model mechanisms can be estimated by averaging performance over all variants that use it.

5.3.2 Model fitting (parameter estimation)

Model fitting refers to the step in which the free parameters (e.g., the utility parameter α in EUT, or the gaze discount parameter θ in the aDDM) of a behavioural model are optimized for the model to best describe the observed choice (and in some cases RT) data. In other words, model fitting is the search for a parameter set that maximises the model-predicted probability (or likelihood) of the observed choice in each trial. In Bayesian parameter estimation (used in Thomas et al., 2019, and by the GLAMbox software package), the *prior* distribution of parameter values is also incorporated, such that parameter values that are unlikely *a priori* are less likely to be estimated.

For a given data set, model fitting yields the set of parameter estimates and an estimate of the likelihood of the data under the model. This likelihood indicates how well the fitted model describes the data and is one criterion on which multiple models are compared (see Model comparison). Parameter estimates from cognitive models often act as interpretable latent variables that are associated with an assumed cognitive mechanism. For example, the estimate of an individual's gaze discount parameter $\hat{\theta}$ in the aDDM quantifies by how much the value representation of unattended alternatives is dampened. In Thomas et al. (2019), these individual parameter estimates were used to investigate individual differences in gaze discounts and their relationship to different metrics of choice behaviour.

5.3.3 Model comparison

Across studies, the comparison and selection between multiple competing models was a central part of the model-based analyses. The goal of model comparisons is to identify the model that described the observed data best and infer from this that people's choices are made similar to its assumed decision process.

One difficulty when comparing computational models of different complexity (e.g., with different numbers of free parameters) is that more complex models can generally fit data better but might also fit features of the data that are not relevant. They *overfit* the data. Added model complexity must therefore be justified by a significant improvement of overall fit. We compared models using the Widely Applicable Information Criterion (WAIC) (Thomas et al. 2019; Vehtari et al., 2017) and Bayesian Information Criterion (BIC) (Molter et al., 2021; Schwarz, 1978), which both take this into account.

In addition to this *relative* type of comparison, it is important to assess different models' predictive performance on an *absolute* level (Heathcote et al., 2015; Palminteri et al., 2017; Wilson & Collins, 2019). To this end, we simulated synthetic data from fitted models and tested whether key patterns in the behavioural data could be reproduced: In Thomas et al. (2019), we tested the two model variants' ability to reproduce individual differences across three behavioural metrics, and the winning model's ability to capture the shapes of individual and aggregate response time distributions. In Molter et al. (2021), we used the different models' predicted association between gaze and choice and their ability to predict individual differences in context effects as an additional *absolute* model comparison criterion.

5.3.4 Individual level of analysis

All computational models were fit to the data of individual participants. First, this circumvents the potential issue that fits to aggregate data might not be representative of the underlying individual processes in which we are ultimately interested (Box 1; Busemeyer & Diederich, 2014; Farrell & Lewandowsky, 2015, 2018). Second, individual model fits are a prerequisite for the model-based analysis of individual differences, which took a central role in Thomas et al. (2019) and Molter et al. (2021).

Likewise, model comparisons were performed to identify the best model for each individual first. Model performance on the aggregate level (e.g., lowest mean BIC across participants) only complemented these findings. This approach was particularly important in Thomas et al. (2019), where we specifically aimed to address whether prior work on the group-level held on the level of individuals, and in Molter et al. (2021), where substantial differences in participants' context-dependent choice behaviour were present.

5.3.5 Validation of modelling analyses

When using computational models for data analysis, either by (i) estimating and interpreting or comparing their parameters (e.g., interpreting gaze discount strength) or by (ii) selecting a best-fitting model from a set of models and inferring that the true data generating process is best described by it, the validity of these claims can be systematically evaluated. To address the first point, it is recommended to test whether a model's parameters can be *recovered* (Heathcote et al., 2015; Lee et al., 2019; Palminteri et al., 2017; Wilson & Collins, 2019). This refers to simulating synthetic data from a model with known parameters, re-fitting the model to the synthetic data set, and testing whether the estimated parameters correspond to the known generating ones. We performed parameter recoveries in multiple studies and further illustrated the steps to perform a parameter recovery using the newly developed model and corresponding toolbox in Molter et al. (2019).

Similarly, model comparison analyses can be validated by performing *model recoveries* (Heathcote et al., 2015; Palminteri et al., 2017; Wilson & Collins, 2019). Here, synthetic data is generated from all competing models, after which each model is fit to each of the synthetic data sets. Finally, model comparison and selection procedures are run, identifying best-fitting models for the data generated from each model.

5.4 Software and data repositories

All statistical analyses including computational modelling analyses, and visualization reported in this thesis were performed in Python (Python Software Foundation) using the numpy (Harris et al., 2020), pandas (McKinney, 2012), PyMC3 (Salvatier et al., 2016), bambi (Capretto et al., 2021), matplotlib (Hunter, 2007), and seaborn (Waskom & seaborn development team, 2020) packages. Data and analysis code for Thomas et al. (2019) are available at https://github.com/glamlab/gaze-bias-differences. GLAMbox (Molter et al., 2019) code is available at https://github.com/glamlab/glambox. GLAMbox documentation including detailed usage examples can be found at https://glambox.readthedocs.io. For Molter et al. (2021), behavioural and eye tracking data, MATLAB (The Mathworks Inc., USA) based task code and Python analysis scripts and models are available at https://github.com/moltaire/gda-context. For Molter and Mohr (2021), jsPsych (de Leeuw, 2015) task code and Python analysis scripts and models are available at https://github.com/moltaire/gaze-choice-causality. The corresponding behavioural task can be run online at https://moltaire.github.io/causality_task.

6 Summary of dissertation studies

In this chapter, I summarise the three empirical studies of the dissertation and briefly highlight the software toolbox developed in the context of the first study.

6.1 Study 1: Investigating individual differences in gaze-biases

In Thomas et al. (2019), we investigated to what extent gaze-dependent evidence accumulation, shown by prior work to describe aggregate choice behaviour well (Krajbich et al., 2010; Krajbich & Rangel, 2011), holds as an adequate model of *individuals'* decision-making processes. This project involved two lines of work:

1. First, a computational modelling framework suitable for application of gaze-dependent accumulation model to data of individual participants had to be developed.

2. Second, this model framework was applied to multiple decision-making data sets to systematically evaluate its performance on the individual level and investigate individual differences in model parameters.

6.1.1 Model development: Gaze-weighted linear accumulator model

In the first step, we developed a computational model serving as an analytical tool to investigate gaze discount effects on the level of the individual. This was necessary, as existing approaches like the aDDM were practically limited in multiple ways: First, because no analytical likelihood of its distribution of first-passage times is available, fitting the aDDM (Figure 1) involved repeated simulation of the accumulation process and comparing the simulated distributions of choices and RTs to the observed empirical data. This procedure is computationally expensive and requires reasonably large amounts of data to obtain stable estimates of the empirical distribution of choices and RTs. Both factors limit the application of this procedure

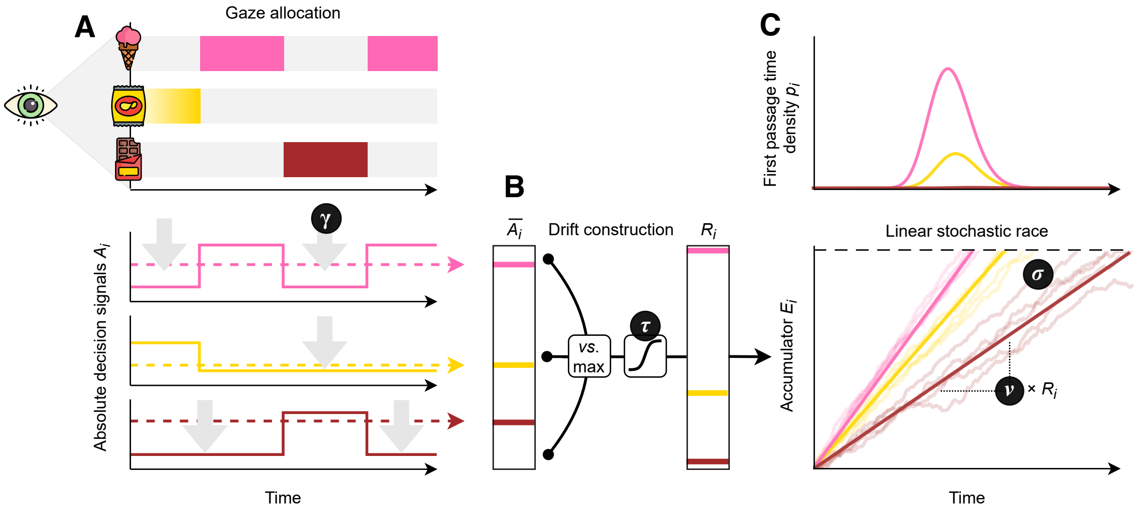


Figure 4. Gaze-weighted linear accumulator model. In the Gaze-weighted linear accumulator model (GLAM), preference construction during the decision process depends on the decision maker's allocation of gaze (**A**). For each item in the choice set, an average absolute decision signal $\bar{A}_i$ is computed (dashed lines in **A**). The magnitude of this signal is determined by the momentary allocation of gaze: While an item is currently not fixated, its signal is discounted by parameter γ ($\gamma \leq 1$; discounting illustrated by gray arrows; **A**). Absolute evidence signals are transformed in two steps to determine a relative decision signal R_i for each item in the (**B**): First, the difference between each average absolute decision signal $\bar{A}_i$ and the maximum of all others $\bar{A}_j$ is taken. Second, the resulting differences are scaled through a logistic transform, as the GLAM assumes an adaptive representation of the relative decision signals that is especially sensitive to differences close to 0 (where the absolute signal for an item is very close to the maximum of all others). The resulting relative decision signals R_i can be used to predict choice and RTs, by determining the speed of the accumulation process in a linear stochastic race (**C**). The stochastic race then provides first-passage time distributions p_i, describing the likelihood of each item being chosen at each time point. Figure and caption adapted from Molter et al. (2019).

to individual level data. Second, since the model's accumulation process depends on visual fixations, the fixation process had to be simulated, too. The development of models predicting fixations' locations and durations is a challenging task by itself, and it grows exponentially more complex for choice settings involving larger choice sets or alternatives with multiple attributes. Additionally, search processes might also differ between individuals, so fixation models would also need to be applicable to each individual.

With the GLAM, we have developed an analytical tool that circumvents these limitations and allows the model-based investigation of the relationship between gaze allocation and choice behaviour at the individual level. It applies to choice situations involving more than two alternatives, and only participants' choices, RTs and relative gaze for each alternative are necessary to apply it, in addition to estimates of the choice alternatives' values. Like the aDDM, the GLAM assumes that the decision process depends on allocation of gaze, as value representations are discounted while the corresponding items are not fixated. The GLAM, however, differs from the aDDM in other important aspects: The fixation-dependent value signals are averaged across the trial by weighting them with the relative amount of time individuals spent fixating the items. This abstracts away the specific sequence of fixations in a trial that are included by the aDDM. On the other hand, this simplification allows for the construction of trial-wise constant drift rates that can enter a basic stochastic race framework.

The framing of the decision process in a race model has two practical advantages: First, race models like the one used by the GLAM often have analytical solutions to their first-passage density distributions, and secondly, they naturally generalize to choice scenarios involving more than two alternatives. The analytical tractability of the race framework further allows for efficient parameter estimation in a hierarchical Bayesian manner. This offers additional advantages like reduced bias, simultaneous estimation of variability, and an intuitive way to quantify uncertainty in parameter estimates (Farrell & Lewandowsky, 2018; Kruschke, 2014; Lee & Wagenmakers, 2013).

To make this tool available to other researchers, we packaged the model code into a Python package called GLAMbox (see Box 1) that can be installed and used without advanced programming knowledge. This toolbox includes additional functionality that goes beyond the model's application in Thomas et al. (2019), like the possibility of estimating model parameters in a hierarchical Bayesian fashion or modeling parameters dependencies on conditions or experimental groups.

Box 1| GLAMbox (Molter et al., 2019)

GLAMbox is a Python-based toolbox that is built upon PyMC3 (Salvatier et al., 2016) and allows the easy application of the GLAM to experimental choice data. The GLAM assumes gaze-dependent evidence accumulation in a linear stochastic race that extends to decision scenarios with many choice alternatives. GLAMbox enables Bayesian parameter estimation of the GLAM for individual, pooled or hierarchical models, provides an easy-to-use interface to predict choice behaviour and visualize choice data, and benefits from all of PyMC3's Bayesian statistical modeling functionality. Further documentation, resources and links to the toolbox itself are available at https://glambox.readthedocs.io.

Feature overview

GLAMbox is published as a Python package on pypi.org (https://pypi.org/project/glambox/). It depends on PyMC3 (Salvatier et al., 2016) for probabilistic programming of the model and parameter estimation, pandas (McKinney, 2012) and numpy (Harris et al., 2020) for data representation and manipulation, and matplotlib (Hunter, 2007) and seaborn (Waskom & seaborn development team, 2020) for visualization. The toolbox and its dependencies can be installed into any Python 3.7 environment with a single command.

```
pip install glambox
```

Listing 6.1. Installing `glambox` and dependencies.

With a formatted data set called `data` at hand, fitting a basic GLAM takes only a few lines of code:

```python
import glambox as gb

model = gb.GLAM(data=data)
model.make_model()
model.fit()
```

Listing 6.2. The basic commands to fit a GLAM to data.

Hierarchical Bayesian parameter estimation

The toolbox supports specification and estimation of hierarchical versions of the GLAM, where individual parameter estimates are assumed to be drawn from a group level distribution. This way, individual parameter estimates are informed by the data from the entire group, exploiting the similarities between participants while not assuming full independence. What is more, a hierarchical model simultaneously models the variance between individuals. This approach can yield less biased parameter estimates and improved parameter estimation especially in the face of limited amounts of data (Farrell & Lewandowsky, 2018; Ratcliff & Childers, 2015; Wiecki et al., 2013). Hierarchical models can be built by setting the model `kind` to `"hierarchical"`:

```python
model.make_model(kind="hierarchical")
```

Listing 6.3. Building a hierarchical GLAM.

Model comparison tools

As in Thomas et al. (2019), it is often necessary to fit and compare multiple competing models. This can be done using the `compare_models` function from the `analysis` module, which returns a data frame with the desired comparison criterion (e.g., WAIC or leave-one-out cross-validation) for each model:

```
from gb.analysis import compare_models
comparison = compare_models(models=[model_1, model_2],
                            ic='WAIC')
```

Listing 6.4. Comparing two GLAM variants.

Dependency setup for comparisons between groups and conditions

Other research questions might require comparison of parameter values between groups or conditions instead of (or complementary to) the comparison of full models. For these situations, the toolbox adapts the `depends_on` keyword from other modelling toolboxes (Wiecki et al., 2013). This keyword allows the user to specify conditional dependencies (both within- and between-subjects) of each model parameter, for individual, pooled and hierarchical models. The model then automatically includes single parameters for each level of the specified condition.

```
model = gb.GLAM(data=data)
model.make_model(kind="hierarchical",
                 depends_on={"gamma": "condition"})
```

Listing 6.5. Using a parameter dependency to compare parameter values between conditions.

Simulation and prediction methods

Model instances have multiple methods that enable fast prediction (simulation) of choice and RT data for a given data set and parameter set. These can for example be used to evaluate absolute model fit, run parameter and model recoveries, or explore model predictions. Below is an example of an out-of-sample prediction routine: The model is fit on a data frame `training`. Then the attached data is exchanged to a `test` data set, and model predictions (50 per trial in `test`) are generated using the parameter values estimated from `training`.

```
model = gb.GLAM(data=training)
model.fit()
model.exchange_data(test)
model.predict(n_repeats=50)
```

Listing 6.6. Example of out-of-sample prediction.

Visualization

The toolbox contains convenient functions to quickly visualize different behavioural measures and their associations, absolute model fit, and posterior parameter distributions. We refer the reader to the examples in the documentation and the original publication for more details.

Publications using `GLAMbox`

- Sepulveda, P., Usher, M., Davies, N., Benson, A. A., Ortoleva, P., & De Martino, B. (2020). Visual attention modulates the integration of goal-relevant evidence and not value. *eLife*, *9*, e60705. https://doi.org/10.7554/eLife.60705

- Thomas, A. W., Molter, F., & Krajbich, I. (2021). Uncovering the computational mechanisms underlying many-alternative choice. *eLife*, *10*, e57012. https://doi.org/10.7554/eLife.57012

- Weilbächer, R. A., Krajbich, I., Rieskamp, J., & Gluth, S. (2021). The influence of visual attention on memory-based preferential choice. *Cognition*, *215*, 104804. https://doi.org/10.1016/j.cognition.2021.104804

- Kaanders, P., Sepulveda, P., Folke, T., Ortoleva, P., & De Martino, B. (2021). Cherry-picking information: Humans actively sample evidence to support prior beliefs. *bioRxiv*. https://doi.org/10.1101/2021.06.29.450332

- Brus, J., Aebersold, H., Grueschow, M., & Polania, R. (2021). Sources of confidence in value-based choice. *PsyArXiv*. https://doi.org/10.31234/osf.io/3wnf7

- Lupkin, S. M., & McGinty, V. B. (2021). Monkey see, monkey choose: A nonhuman primate model of gaze biases in economic choice. *Poster presented at the Society for Neuroscience Global Connectome Virtual Meeting*

6.1.2 Empirical analysis of individual differences in gaze-dependent evidence accumulation

With the GLAM developed, we set out to empirically test the extent of individual differences in gaze-dependent evidence accumulation processes across four existing data sets spanning two- and three-alternative choices in value-based and perceptual choice domains (Krajbich et al., 2010; Krajbich and Rangel, 2011; Experiment 2 from Folke et al., 2016; Experiment 1 from Tavares et al., 2017).

First, we investigated individual differences in key measures of decision making: The probability of choosing the best item in a trial, the time it took to make the choice, and the estimated influence of gaze on choice. This last measure represents the average increase in choice probability for an item that is looked at longer than others in a trial, when value is taken into account and was devised as a model-free descriptor of an individual's association between gaze and choice.

Next, the GLAM and a variant without gaze-dependence were fitted to the data of each individual using Bayesian parameter estimation. This allowed us to perform a principled model comparison on an individual level and address the question of whether the finding that group-level data are well-described by gaze-dependent accumulation holds on the individual level. We found that 109 out of 118 (92%) participants were better described by the model with gaze-dependence, providing strong empirical evidence that a gaze bias mechanism is present for most individuals across tasks and choice domains. This analysis also revealed substantial individual differences in the gaze-discount parameter γ in each data set, such that unattended values were strongly discounted for some participants but much less affected by gaze for others.

We then analysed the absolute fit of the model to the data by testing whether the model could accurately predict individual differences across the three behavioural metrics and generally capture the shape of individual and aggregate response time distributions. We found that the model with gaze-discount predicted individual differences in all metrics, within and across data sets and that a model without gaze-discount failed to account for the association between gaze and choice, as expected.

Finally, we tested whether model parameters could be used to predict individual differences in behaviour. Our analyses suggest that individual differences in gaze-discount strength are associated with individual differences in the ability to choose items that are valued higher than others when evaluated individually.

In summary, leveraging the developed GLAM framework and toolbox (Box 1), we found strong support that gaze-dependent accumulation is a good model of individuals' choice behaviour across decision-making domains and choice set sizes (two vs. three alternatives): It captures key characteristics like RT, gaze-influence on choice, the choice itself and, crucially, individual differences in these measures, while models without gaze-dependence fail to account for the observed relationship between gaze and choice. Importantly, we identified gaze-discount strength as a potential source of sub-optimal choice behaviour.

6.2 Study 2: Gaze-dependent accumulation in context-dependent risky choice

Thomas et al. (2019) confirmed that gaze-dependent evidence accumulation provides a good characterisation of individual choice behaviour in simple decision-making tasks. In Molter et al. (2021) we aimed to further investigate whether this finding holds, even when choices are influenced by the context of available alternatives as in the presence of context effects in risky choice.

To this end, we recorded participants' eye movements while they made choices between three risky gambles, each described by a probability p to win an amount m and nothing otherwise. This multi-alternative, multi-attribute choice task on its own presented a more complex decision setting than those tasks analysed in Thomas et al. (2019). Furthermore, however, choice sets were designed to elicit compromise

(Simonson, 1989) and attraction effects (Huber et al., 1982) known to violate axioms of rational choice, posing a significant challenge to models of preferential choice.

In line with previous findings, we found choice behaviour to be context-dependent: Choices were influenced by the context of available alternatives, and participants showed both compromise and attraction effects. The degree to which participants exhibited these effects, however, was subject to large interindividual differences, with some participants preferring dominant alternatives in trials designed to elicit attraction effects, particularly frequently. These behavioural results provided a complex testing scenario to compare the gaze-dependent accumulation with competing theoretical accounts of both standard risky and context-dependent multi-alternative, multi-attribute choice.

Importantly, and a prerequisite for a simple gaze-dependent evidence accumulation model to capture context effects in participants' choices, we found that gaze allocation was also modulated by the context of available alternatives such that dominant and compromise alternatives received longer relative gaze during the course of the decision.

This allowed a simple gaze-dependent evidence accumulation model derived from prior work on binary risky choice (Glickman et al., 2019) to outperform both established models of risky (EUT; Von Neumann & Morgenstern, 1947) and context-dependent multi-alternative multi-attribute choice (MDFT; Roe et al., 2001), and provide the best description of choice data. In addition, this model quantitatively captured the association between gaze (specifically, the relative amount of time an alternative was looked at more than others) and choice. We found, however, that the model underestimated particularly strong attraction effects because eye movements were distributed more evenly between alternatives than choices (e.g., dominated decoys were rarely chosen but still looked at for a significant portion of deliberation), limiting the model's predictive range.

Finally, we performed a systematic search across a large space of possible model variants, combining single mechanisms of multiple model classes, and exploring additional mechanisms of gaze-dependence (e.g., gaze-dependent accumulation leak and inhibition). In this switchboard analysis (cf. Turner et al., 2018) we showed that all participants' behaviour was still best described by some form of gaze-dependent accumulation and that the variant describing most participants best coincided with the *a priori* defined, winning model. Additionally, however, this analysis revealed that predicting data from participants with particularly strong attraction effects required an additional similarity-based inhibition mechanism.

In summary, this study provided additional support of gaze-dependent evidence accumulation as a framework of value-based choice, even in risky choices that are more complex than simple choice between multiple snack foods, and even when choices are context-dependent as in attraction and compromise effects, although the basic model needed to be extended to capture extreme effects.

6.3 Study 3: Causal effects of presentation duration and order on binary risky choice

Thomas et al. (2019) and Molter et al. (2021) generally confirmed positive associations between gaze allocation and choice behaviour and demonstrated that gaze-dependent evidence accumulation can provide precise descriptions of these variables' interactions, even on the individual level and in the presence of contextual influences on choice.

A remaining concern, which we addressed in Molter and Mohr (2021), is the question of causality: Do aspects of information search like gaze duration causally affect people's choices, or do they rather only reflect emerging preferences? As gaze-dependent computational models of choice (including GLAM and models used in Molter et al., 2021 do not imply a causal direction in the association between gaze and choice, and answering this question ultimately requires experimental manipulation of decision makers' information search process.

In addition to the duration for which alternatives are inspected, other aspects of information search have been associated with choice behaviour: There are, for example, multiple findings relating the *order* in which information is acquired to choice. The gaze cascade effect describes the finding that chosen alternatives are attended increasingly over the course of deliberation. Similarly, chosen alternatives are typically fixated last, just before a choice is indicated. Furthermore, the winning model from Molter et al. (2021) predicted an effect of presentation order, as it included a leak mechanism whereby information acquired later in the decision is weighted more heavily than earlier acquired information.

With viewing duration and order both linked to choice, one potential pitfall in the study of these individual factors is that they can be easily confounded. This might be especially true in simple, fast, and repeated choices often studied in the laboratory (e.g., as in the data sets analysed in Thomas et al., 2019), where decision makers make only a few fixations before their choice.

Here, we therefore investigated independent causal effects of viewing duration and order on decision making in a task involving choices between two risky gambles, each described by a winning probability p to win an amount m and nothing otherwise (as in Molter et al., 2021). Importantly, information about the two gambles' attributes was presented sequentially, allowing the precise control of presentation duration and order. Our task further included alternative-wise and attribute-wise presentation formats.

All data collection and analysis procedures were fully preregistered (Molter & Mohr, 2021a).

Against our hypotheses derived from causal interpretations of gaze-dependent accumulation models, we found evidence against causal effects of presentation duration on choice in this task (both in attribute- and alternative-wise presentation formats). In contrast, we found presentation order to causally affect choices, such that alternatives that were shown last before a choice was prompted were more likely to be chosen. Evidence also favoured an order effect in attribute-wise presentation, such that alternatives with better values on the last-presented attribute were chosen more frequently.

7 General discussion

In this final chapter, I will begin by first discussing how the three studies addressed the research questions. After this, the studies' results will be discussed more generally in the broader context of constructed preferences and computational modelling of decision making. To this end, I will embed the thesis' results into an updated version of the framework of decision making introduced earlier.

7.1 Discussion of the research questions

Question 1. *How can individual gaze-bias mechanisms be investigated efficiently using computational modelling?* In Thomas et al. (2019) and Molter et al. (2019), we developed a novel gaze-dependent evidence accumulation model inspired by a multi-alternative version of the aDDM (Krajbich & Rangel, 2011), that applies to choice scenarios involving an arbitrary number of choice alternatives, while remaining analytically tractable. It enables estimation of gaze discount parameters on the individual level and advanced modelling techniques like Bayesian hierarchical parameter estimation.

The GLAM offers a way to circumvent the complex requirement of other gaze-dependent modelling approaches (like the aDDM) to build generative models of the visual search process. Although researchers have started to explore such models in simple decision-making tasks and to integrate them with accumulation-based models of choice (Callaway et al., 2021; Gluth et al., 2020; Jang et al., 2020; Towal et al., 2013), these solutions remain difficult to apply to new tasks. Also, specific fixation sequences might not be of main interest to the researcher but rather their aggregate association with choice. Here, the GLAM provides a tractable, but simplified, alternative to the aDDM.

Other approaches using similar forms of simplification in favour of computational tractability exist for the case of two-alternative choice (e.g., Cavanagh et al., 2014;

Smith et al., 2019). First applications of the GLAM by other researchers enabled by our Python toolbox (Box 1), however, highlight another advantage associated with its race architecture (Brus et al., 2021; Sepulveda et al., 2020): The difference between accumulated evidence in favour of each alternative at the time of choice is often construed as a measure of decision confidence (De Martino et al., 2013; Vickers, 1979). Traditional diffusion models (including the DDM and aDDM) use a *relative* choice rule, where the relative evidence of one alternative over the other is accumulated until a threshold is reached. This, however, implies that the evidence difference at time of choice is constant, and cannot be associated with variance in confidence judgments. Notably, however, while the GLAM's accumulation process occurs independently for each item, the drifts are constructed comparatively, combining elements of fully independent race models (e.g., Vickers, 1970) , and comparative accumulation models (e.g., Krajbich & Rangel, 2011; Ratcliff, 1978; Roe et al., 2001). This way, Brus et al. (2021) and Sepulveda et al. (2020) were able to investigate associations of gaze, choice, and confidence using the GLAM.

In sum, our work demonstrated a way to model individual gaze bias effects efficiently by aggregating within-trial dynamics and embedding a gaze-discount mechanism into a computationally tractable framework.

Question 2. *Does gaze-dependent evidence accumulation capture simple choice behaviour on the individual level?* Extending prior work which reported gaze-dependent evidence accumulation to accurately describe *aggregate* choice behaviour (Krajbich et al., 2010; Krajbich & Rangel, 2011), our different lines of work established gaze-dependent evidence accumulation as a general model of decision making *on the level of the individual*: In Thomas et al. (2019), we found that across four data sets spanning two- and three-alternative, perceptual, and value-based (snack food) choices, collected in different laboratories, almost all individuals were best described by GLAM variants with gaze bias mechanisms. Importantly, the model also accurately predicted individual choice behaviour across multiple metrics on an absolute scale. Similarly, in Molter et al. (2021), where we investigated choices between three risky gambles, all participants' choice behaviour was best described by models that included forms of gaze-dependence. In recent related work we could show gaze-dependent accumulation to also capture individual choices, response times, and associations with information search in even larger choice sets with up to 36 snack food items (Thomas et al., 2021).

Across studies, we demonstrated that models without gaze-dependence fail to capture the robust associations between gaze and choice that are observed empirically, challenging existing models of decision making that do not include links to visual attention.

There are, however, limitations to the applicability of basic gaze-duration-dependent models like the aDDM or GLAM, as additional mechanisms were necessary to explain some patterns of choices in Molter et al. (2021), and no effect of presentation duration on choice could be identified in Molter and Mohr (2021; see below). Furthermore, Smith and Krajbich (2018) found that a group of participants in a social decision-making task rather followed a simple choice rule where they searched for the alternative with the highest outcome for themselves, inconsistent with gaze-dependent evidence accumulation. The specific conditions under which decision makers' choice behaviour follow gaze-dependent accumulation versus when other processes or strategies are used and how decision makers arbitrate between strategies, remain important questions for future work.

In sum, our findings provide strong evidence for group-to-individual generalizability (A. J. Fisher et al., 2018) across decision domains and set sizes, and thereby an increased understanding of *individual* decision making. Assessing mechanisms like the gaze discount on the individual level furthermore lays the foundation for the model-based analyses of individual differences in decision making discussed in the next section.

Question 3. *Which interindividual differences exist in gaze-bias strength and how do they relate to individual differences in choice behaviour?* Across studies, we found that the large majority of people show a positive association between gaze and choice such that alternatives that are looked at longer are also more likely to be chosen. Importantly, however, individual strengths of this gaze bias varied substantially, with some participants' choices being strongly associated with their gaze allocation, whereas the association was weak for others.

Individual estimates of gaze discount parameters in Thomas et al. (2019) reliably predicted decision makers' ability to choose the best item from a choice set, suggesting the gaze discount mechanism as a source of individual differences in the ability to choose consistently.

Our findings thereby suggest that individual gaze bias strength — and the gaze discount as its underlying mechanism — are meaningful measures of individual differences in decision making whose association with other individual measures should be investigated further. Initial work found that gaze bias strength can be associated with psychophysical metrics like the degree of "tunnel vision", that is, the spatial extent of a person's visual attentional scope (Smith & Krajbich, 2018). This view implies that gaze discount strength can be considered a trait that is constant for each person across decision scenarios. Other work (Brus et al., 2021) building on the GLAM framework, found that gaze discounts should be construed as variable between trials. The relative contribution of variability between and within individuals will need to be addressed in the future.

The model-based analysis of behaviour and individual differences therein has also seen increasing application in clinical settings (Huys et al., 2021; Huys et al., 2016): Computational psychiatry uses algorithmic models of behaviour (like the DDM or reinforcement learning models) to test mechanistic hypotheses about the processes underlying psychological disorders. In addition, computational models are useful measurement devices, which enable the measurement of hidden variables associated with distinct psychological mechanisms. Resulting estimates can be compared across groups or related to individual symptoms (Huys et al., 2016). Finally, an increased understanding of the generative processes underlying disease can help motivate novel therapeutic approaches and interventions (Huys et al., 2021; Huys et al., 2016).

In this context, Vaidya and Fellows (2015) found that patients with frontal-lobe-lesions exhibited less consistent decision making in a preferential choice task. Using computational modelling, they found that patients had steeper gaze discounts than controls. The use of the computational model in this case provided an improved understanding about the functional origin of the observed behavioural change, namely the exaggerated discounting of unattended information.

In a similar vein, lifespan differences in decision making might be associated with differences in information search and gaze biases: Older adults often search for less information and use different information search strategies (M. M. S. Johnson, 1990). Diminished working memory capacity has been suggested as a potential cause (Mather, 2006), and could possibly be related to gaze discount mechanisms that also act on momentarily unattended but memorized information. In a meta-analysis of older adults' risky decision making Mata et al. (2011) found that inconsistent results on older adults' differences in risk attitudes could be explained by different choice tasks' learning requirements. Similarly, the way information must be searched in different tasks or ageing-related changes in gaze bias strengths (or both) could explain behavioural differences.

To understand adolescents' risk preferences, Ciranka and van den Bos (2021) argue for a more ecological perspective which emphasizes adolescents' environments and exploration behaviour. As the deployment of visual attention constitutes a core aspect of exploration behaviour, this also suggests the potential relevance of gaze bias mechanisms in understanding adolescents' decision behaviour. Experimental evidence for an association of adolescents' search and choice behaviour comes from Kwak et al. (2015) who find adolescents' choices to be more risk averse compared to young adults, with eye movements showing more systematic, analytic information search. Notably, this result conflicts with the typical observation of adolescents acting more impulsively (Blakemore & Robbins, 2012). Here, the analysis of adolescents' decision behaviour using gaze-dependent computational models of decision making might provide novel insight into differences in adolescents' preference construction. In this regard, recent work has demonstrated links between measures of attention and risk preferences characterized in PT' parameters (Pachur et al., 2018; Zilker & Pachur, 2021). The computational nature of the GLAM might further help move the field of developmental cognitive neuroscience towards more specific and testable theories (van den Bos & Eppinger, 2015).

In sum, our work revealed large interindividual differences in gaze discounts, which were associated with individual differences in choice consistency and can inform potential future clinical and lifespan decision-making research.

Question 4. *To what extent can gaze-dependent evidence accumulation explain complex multi-attribute, multi-alternative risky choice?*

Molter et al. (2021) showed that gaze-dependent evidence accumulation models outperformed both traditional accounts of risky choice (EUT) and an established dynamic cognitive model of context-dependent choice (MDFT) in a task setting where participants' choices between three risky gambles were systematically influenced by the context of available alternatives. The model with the best fit to the data was a straightforward extension of a model proposed for two-alternative choice (Glickman et al., 2019), which assumes leaky accumulation of alternative-wise subjective values, with a simple gaze discount, similar to the aDDM and GLAM. As in Thomas et al. (2019), only gaze-dependent models were able to account for the positive association of gaze duration and choice. The fact that prediction of particularly strong attraction effects required inclusion of an additional similarity-dependent inhibition mechanism, however, highlights possible limitations of the most basic forms of gaze-dependent discounting and accumulation leakage. Yet, even in these cases, gaze-dependence remained an important feature of the model.

Notably, in contrast to most other cognitive models of context-dependent choice able to predict context effects (e.g., Roe et al., 2001; Tversky, 1972; Tversky & Simonson, 1993; Usher & McClelland, 2004), the best-performing model in Molter et al. (2021) employed alternative-wise valuation (i.e., integration of an alternative's winning probability and amount attributes), not comparisons between multiple alternatives' attribute values within single attribute dimensions. Prior work has argued that these attribute comparisons are essential to predicting context effects (Noguchi & Stewart, 2014). Our results demonstrate that context effects can also emerge from models with alternative-wise valuation in conjunction with gaze-dependence, as decision makers' gaze allocation itself is context-dependent.

Taken together, we found gaze-dependent evidence accumulation to generalize to multi-attribute, multi-alternative risky choice, and gaze-dependence as a prerequisite to account for the association of gaze and choice. Our results, however, also illustrate potential limitations of gaze-dependent accumulation in its most basic forms.

Question 5. *Can binary risky choices be causally influenced by external control of presentation duration?* In contrast to our predictions and causal interpretations of gaze-dependent accumulation models we could not replicate a causal effect of presentation duration on choice reported in prior work (Armel et al., 2008; Shimojo et al., 2003) in Molter and Mohr (2021).

Prior work already found effects of experimentally manipulating viewing or presentation duration on choice to be smaller than reported associations in free viewing tasks (Krajbich, 2019). There are multiple possible explanations for this: First, in paradigms where presentation duration is controlled, presentation durations might not translate directly into viewing durations, for example, due to decision makers needing to shift their gaze synchronously to the presentation stream, or them not fixating only information on the screen. Second, viewing the presented information does not necessarily imply conscious, attentive processing (Orquin &

Holmqvist, 2018), which could be required to obtain influencing effects. Externally controlled presentation paradigms in particular — necessarily — show stimulus information either for shorter or longer than the time participants would have taken to make their choice in a self-paced and free-viewing setting. Therefore, participants might have implicitly made their decisions before they were prompted to indicate their choice, preventing the full presentation sequence and duration differences from affecting their choice process. Third, decision makers could rely on peripheral vision and not orient their gaze towards the presented information. Shimojo et al. (2003) reported these orienting movements to be required to observe causal duration effects. Finally, a bi-directional association between gaze and choice in decision making with natural viewing, as postulated by the gaze cascade theory and described in recent work (Gluth et al., 2020), would also result in a weaker association when information presentation is controlled and preferential viewing thereby suppressed. Designs using gaze-contingent choice prompts can ameliorate some, but not all of these drawbacks: On the one hand, while decision makers can inspect information freely, they might still be prompted to make a choice only after they already committed to an alternative implicitly. On the other hand, the conditions triggering the choice prompt (e.g., difference in viewing duration between alternatives) might never be fulfilled in a trial. Such "ineffective trials" can further induce artificial effects if they are removed from analyses (Ghaffari & Fiedler, 2018; Newell & Le Pelley, 2018)

Investigating the causal contribution of aspects of the information search process on choice, therefore, remains challenging, as experimental manipulations often interfere with participants' natural decision-making routines. One recently used approach provides an elegant solution to reducing this interference: Gwinn et al. (2019) used a separate visual search task to induce an attention bias towards one screen side, which was carried over to the main choice task. This way, participants could freely search for information and indicate their choice naturally, without interference by the experimenters. Importantly, the authors found the attentional manipulation to produce downstream effects on choice. Notably, this effect was mediated by the location of the first fixation and not viewing duration, suggesting an effect of acquisition *order*.

In sum, unlike prior work (Armel et al., 2008; Liu, Lyu, et al., 2020; Pärnamets et al., 2015; Shimojo et al., 2003; Sui et al., 2020; Tavares et al., 2017), we did not find a causal effect of presentation duration on choice, suggesting that its presence is not universal. Possible factors moderating this effect need to be investigated in future work, which will also allow a better judgment over the relevance of these effects in real-world scenarios.

Question 6. *Can binary risky choices be causally influenced by external control of presentation order?* In Study 3 (Molter & Mohr, 2021b), instead of the predicted effect of presentation duration, we found strong evidence in favour of an effect of presentation *order* on binary risky choice behaviour, such that alternatives presented last, just before a choice was prompted, were chosen more frequently. Our data also suggested an order effect on the level of attribute dimensions, where alternatives with better values on the attribute dimension shown last were chosen more frequently. This effect is consistent with models of gaze-dependent evidence accumulation that use a form of accumulation leak (like the winning model in Molter et al. (2021), where changes to the order of information result in changes to the resulting choices, as later acquired information is weighted more heavily relative to information acquired early. Interestingly, this kind of recency effect in decision making has been demonstrated in decisions-from-experience and related paradigms, where decision makers repeatedly sample outcomes from risky prospects and have to learn the outcomes' probabilities (Hertwig et al., 2004; Tsetsos et al., 2012). In these paradigms, decision makers also encounter information in a sequential fashion. Our results indicate that a similar, causally directed effect is also present in decisions from description, where information is also encountered sequentially, including settings when the decision maker seeks information by herself.

Furthermore, our results can be linked to literature on decisions from memory (Gluth et al., 2015; Weilbächer et al., 2021). There, a memory bias can be found, where better remembered alternatives are chosen more frequently across trials, even though their value is below average. The recency effect we find can be interpreted similarly, as the last presented information is least likely to be forgotten or confused. If participants had a bias for alternatives that they remember, this recency effect would be expected.

Finally, together with the evidence against a causal effect of presentation duration on choice, we see the possibility that some prior work might have confounded duration and order effects, as chosen alternatives are typically looked at last *and also* longer. This pattern might be particularly prevalent in fast-paced, repeated decisions in an experimental setting, where viewing patterns like "chosen-unchosen-chosen" are common.

Taken together, we identified a causal effect of presentation order, in the direction of a recency effect, in binary risky choice. Our results highlight the importance of controlling and differentiating causal duration and order effects.

7.2 A constructive role of visual attention in decision making

All studies contained in this thesis contribute to a broad body of research on the *constructive* nature of preferential choice, with the core assumption that observed preferences are not the results from referencing a "master list [of values] in memory" (Payne et al., 1992, p. 89), but actively constructed by the decision maker's interaction with the choice problem and its context:

In Thomas et al. (2019) and Molter et al. (2021), we found choice, RT, and gaze data to be well described by a gaze-dependent evidence accumulation model, where choices are constructed in a dynamic context-dependent process. Notably, emphasizing its constructive character, the model was also able to account for individuals' RT distributions, that is, predict how long decision makers would take to make their choice, on a trial-to-trial basis (Thomas et al., 2019).

Assuming an active role of visual attention in the decision process (e.g., in the form of a gaze discount mechanism) can account for both description- and procedure invariances, as differences in the presentation of choice alternatives or the task procedure (e.g., choosing vs. pricing) can be associated with changes to decision makers' gaze allocation (Kim et al., 2012; Orquin & Mueller Loose, 2013). Furthermore, as shown in Molter et al. (2021), the context-dependent allocation of gaze can result in context-dependent and, therefore, constructed choice behaviour.

In the classification of explanations by Payne et al. (1992), gaze-dependent evidence accumulation generally provides a *perceptual* explanation of constructed preference phenomena. As Payne et al. (1992) notice, however, integration with cost/benefit oriented frameworks, specifically with regard to the adaptive selection of strategies, is possible. To this end, future work might explore and explicitly model different strategies of attention allocation (see below) or different quantities (e.g., ordinal comparisons instead of alternative-wise values; Noguchi & Stewart, 2018) that are integrated in a gaze-dependent fashion.

One concern with these explanations is the question of causality. If gaze merely reflected other constructive processes during choice, no additional insight would be gained by using it to explain choice. However, together with prior work demonstrating causal influences of viewing duration and other aspects of information search (Armel et al., 2008; Gwinn et al., 2019; Liu, Lyu, et al., 2020; Liu, Zhou, et al., 2020; Pärnamets et al., 2015; Shimojo et al., 2003; Sui et al., 2020; Tavares et al., 2017), Molter and Mohr (2021) refutes this account of "epiphenomenal" gaze, as we found evidence for a causal influence of presentation order on choice.

This order effect highlights that another temporal dimension of information search can cause description invariance, which goes beyond differences in relative gaze duration: Factors which affect the temporal order of information acquisition (e.g., when information is delivered in a controlled sequential way, as in video advertisements; or a decision maker's tendency to search information in a certain way) are actively involved in the decision maker's preference construction process.

The gaze-dependent account of preferential choice applied in this thesis falls between two different perspectives on the constructive roles of information acquisition and use, taken previously: On the one hand, judgment and decision-making phenomena have been attributed to congruent biases in the sampling or use of information within the decision maker (e.g., increased sensitivity to losses by PT's asymmetric value function; Weber & Johnson, 2009). On the other hand, the more ecological perspective on judgment and decision making by K. Fiedler (2000) assumes behavioural biases to emerge from sampling biases *caused by environment*. Since visual attention in decision making is influenced both by factors of the environment (e.g., the context of available alternatives), as well as factors internal to the decision maker (Corbetta & Shulman, 2002; Orquin & Mueller Loose, 2013), gaze-dependent evidence accumulation models of choice lie between these two perspectives.

Our work differs from earlier literature that also attributed behavioural evidence of constructed preference to "attention", often taking the role of information weights (for a review see Weber & Johnson, 2009), in that it instead directly refers to eye movement data. The use of "attention" as a general explanatory device has recently been criticized (B. Anderson, 2011; Hommel et al., 2019; Krauzlis et al., 2021), in part to the many meanings it can take and the risk of circular arguments. Instead, focusing on eye movement recordings reduces such ambiguities and, importantly, establishes a clear connection to operationalized and observable process data.

This thesis' results complement recent other work reporting evidence accumulation informed by eye movements to account for different decision making patterns: Gluth et al. (2020) and Gluth et al. (2018) demonstrated that models derived from the multi-alternative aDDM (Krajbich & Rangel, 2011) can predict certain distractor effects better than two competing accounts, when the choice alternatives' values can affect allocation of visual attention. In the context of risky choice, Zilker and Pachur

(2021) found that attentional biases in the aDDM are reflected in PT's probability weighting parameters, suggesting that behavioural effects typically associated with and explained by PT's non-linear probability weighting might also be the effect of simple attentional biases (see also Pachur et al., 2018).

7.2.1 Visual attention and limited resources accounts

The analysis of visual attention during the decision process is also linked to the theoretical perspectives of bounded (Simon, 1955), or resource rationality (Lieder & Griffiths, 2020), which stress inherent constraints on human cognition: Decision makers acquire information about available choice alternatives, their attributes, and associated actions (e.g., derived from different alternatives' position on the screen) *sequentially*, and do not extract all of this information from the environment instantaneously. Similarly, and in contrast to theories of rational choice, not all available information is acquired every time (e.g., some alternatives are not fixated in large choice sets; Thomas et al., 2021). Our work confirmed these limitations on information search to be important components of the decision-making process. The kind of gaze-dependent evidence accumulation models used in our studies formally include this sequential property of information search and can thereby form the basis for resource-rational analyses (Lieder & Griffiths, 2020) of decision making, as performed recently (Callaway et al., 2021). In this study, the authors used the aDDM and, crucially, approximated the optimal policy in which agents *should* allocate their gaze given the constrains by the model's gaze-dependence and the cognitive costs associated with shifting gaze. This way, they found that attending alternatives whose value estimates are high and uncertain, which aligns with empirical data in two- and three-alternative simple choice (Gluth et al., 2020; Krajbich et al., 2010; Krajbich & Rangel, 2011).

7.3 Dwelling on computational modelling of choice

Given this thesis' focus on computational modelling of choice as an analytical method, and theory of the decision-making process, this section will summarize implications for future model- and (viewing models as formally specified theories; J. G. Johnson & Frame, 2019; Lewandowsky & Farrell, 2010) theory building, and discuss methodological issues and possible solutions.

The most pervasive finding of this thesis' studies with respect to the behavioural modelling of choice is the significance of incorporating a modulating effect of gaze on the choice process. Across tasks and data sets, models that make this assumption outperformed competing accounts in predicting choices (and RTs), and were the only models able to account for observed relationships between gaze duration and choice. Conversely, its omission repeatedly showed to miss robust features of the data.

In addition to explaining differences in behaviour between participants (Thomas et al., 2019), gaze-dependence can account for variability within individuals, across occasions: Human choices are inherently stochastic (Luce, 1959; Nesselroade & Ram, 2004; Rieskamp, 2008) so that decision makers do not make perfectly identical choices across otherwise similar occasions. This "noise" (Kahneman et al., 2021) is often modelled using probabilistic choice rules (Luce, 1959; Sutton & Barto, 2018) without clear process interpretations, or assuming variance in hypothetical utility signals (McFadden, 1973). Here, choice models including explicit gaze-dependence allow part of this choice variability to be attributed to the allocation of gaze, which can be scrutinized further.

A second feature that Molter et al. (2021) and Molter and Mohr (2021) suggest to be an important feature of evidence accumulation models of decision making is a form of *accumulation leak* (Usher & McClelland, 2001) or similar mechanisms able to reproduce the causal effect of the order in which information is considered. Future iterations of the GLAM will likely benefit from such an extension, even though adding a leak term to it is not trivial, as it was explicitly designed to average across the temporal dynamics within a trial. Including serial position into the gaze-weighting mechanism could be one possible solution.

Given the constructive role of gaze on the decision process, the question of what drives gaze during deliberation gains new relevance. Numerous studies have addressed this question in a wide range of settings (for a review see Orquin & Mueller Loose, 2013), including risky choice (S. Fiedler & Glöckner, 2012; Glöckner & Herbold, 2011; Stewart et al., 2016). While these insights are sometimes used to inform theory or model *development* (Noguchi & Stewart, 2014), a more frequent use for them is to aid *selection* between extant models. This way, process data have been interpreted to either support, or contradict different decision-making theories by comparing them with theoretically inferred predictions (e.g., S. Fiedler & Glöckner, 2012; Glöckner & Herbold, 2011; E. J. Johnson et al., 2008; Noguchi & Stewart, 2014; Orquin & Mueller Loose, 2013; Stewart et al., 2016). Process data appearing to be compatible with a theory, however, remain difficult to use as a criterion of model selection for multiple reasons: First, there might be other compatible models that were not considered. This highlights the importance of selecting an appropriate *model space*, as, trivially, only models included in the model space can be declared the best account of the data. In Molter et al. (2021), we therefore adopted an approach to systematically generate a large space of possible models, which included variants of competing accounts (i.e., one variant resembled MDFT in many details). Please note, however, that this large model space is also far from exhaustive, as it did not, for example, cover heuristic models of choice.

Second, the important step of model falsification (see Palminteri et al., 2017) using process data can be limited too when models do not make clear predictions about the recorded process data. EUT or PT, for example, are agnostic about the choice *process* and do not specify which eye movements should be made during choice. It is only with additional assumptions (e.g., E. J. Johnson et al., 2008) that those predictions are derived. Many decision making models, for example, predict weights with which a given attribute (e.g., an alternative's quality) contributes to choice (e.g. MDFT, Roe et al., 2001). Does a lack of correlation between gaze duration towards this attribute and the model-predicted weight imply that the model should be rejected? No, because the model does not specify a relationship between its attribute weight and gaze data. It could still be that decision weights are rather reflected in fixation counts (or, alternatively, not in any type of eye movement data at all). As long as the model is not specific about its components' relationship to observable types of process data, these inferences remain difficult to make confidently.

There are multiple ways to increase the link between decision-making models and process data like eye movement recordings, and thereby allow stronger inferences based on the process data: The first, which Thomas et al. (2019) and Molter et al. (2021) followed, is the direct input (Turner et al., 2017) of process data into the decision model. This way, models made quantitative and testable predictions about the relationship between gaze and choice data.

Molter et al. (2021) further show that the inclusion of gaze-data also acts as a constraint for the models' predictions, as the simpler gaze-dependent model (without additional mechanisms) predicted choices of dominated decoy alternatives too frequently, due to distribution of gaze between alternatives not being as extreme as choices (decoys were looked at more than they were chosen).

A more involved, second way is simultaneous modelling (Turner et al., 2017) of both the choice process *and* the way in which process data are generated. In the context of this thesis, this means not only having a gaze-dependent evidence accumulation process but also including a theory on how gaze is allocated. In this case, process data can still be used to evaluate the joint model, by comparing either the observed process- or choice data to those predicted by the model.

Crucially, this approach represents a way to integrate decision making theory with theories concerning the allocation of visual attention (Itti & Koch, 2000, 2001; Orquin & Mueller Loose, 2013), and helps understand how different drivers of gaze affect decision making. For example, we found gaze to be allocated in a context-dependent manner in Molter et al. (2021), yet the question of *why* it is distributed this way remains to be addressed. While initial work using the aDDM included basic fixation-generating models assuming positional effects on gaze (e.g., an initial bias towards the left side; Krajbich et al., 2010; Krajbich & Rangel, 2011), more recent accounts revealed influences of visual saliency (Towal et al., 2013), option value (Gluth et al., 2018), and momentary level of favourable evidence (Gluth et al., 2020) on gaze allocation (and thereby indirectly on choice) using simultaneous modelling.

7.4 Contributions in the framework of decision making

The results of our studies integrate into and extend the framework of decision making presented earlier (Figure 3), emphasizing the constructive role of visual attention in the decision-making process:

Processes of *representation* were not specifically addressed by this thesis' studies. It can be argued, however, that the consideration of decision makers' visual attention provides a direct and measurable way to assess their representations of the choice set. A recent study related to the work presented here supports the notion that decision makers use visual attention to select and represent only a subset of all available items for further consideration: Thomas et al. (2021) found gaze-dependent evidence accumulation successfully predicted choices between snack foods in large choice sets with up to 36 items. Notably, decision makers did not fixate all items. The GLAM variant which best predicted behaviour included only fixated items into the decision process, thereby effectively implementing a gaze-dependent formation of consideration sets.

While not the main focus of this thesis, our studies showed models which assume alternative-wise valuation to provide the best descriptions of choice behaviour. For the choice between risky prospects specifically, in line with prior work (Glickman et al., 2019) valuation by within-alternative integration of attributes was favoured

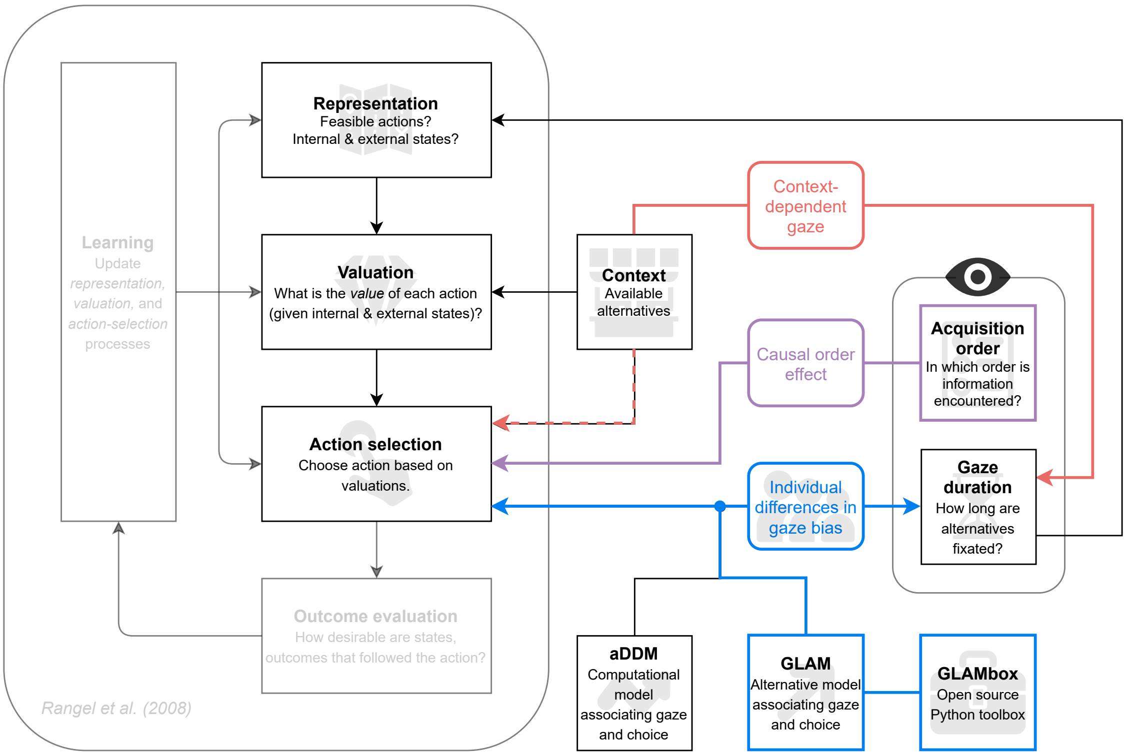


Figure 5. A refined role of visual attention in the framework of decision making. The studies from this thesis have further extended and refined the framework of decision making (Figure 3), particularly the role of visual attention in action selection, in multiple ways. In Thomas et al. (2019) (blue), we confirmed the generally positive association between gaze duration and action selection. Using our novel GLAM framework and toolbox, we confirmed gaze-dependent evidence accumulation to capture *individual* choice behaviour. We revealed, however, large interindividual differences in gaze discounts. Molter et al. (2021) (red) similarly confirmed the positive association between gaze duration and choice. In addition, we established a novel possible path in which the context of available alternatives affects the decision-making process, namely mediated by gaze allocation. In Molter and Mohr (2021) (purple), we probed the causal direction between action selection and information search, specifically presentation duration and order of stimulus information. We found acquisition order to causally affect choices, as last-shown alternatives were chosen more frequently.

over attribute-wise comparisons in Molter et al. (2021). It remains, however, debated to what extent decision makers perform alternative-wise *valuation* (Hayden & Niv, 2021; Vlaev et al., 2011) or select actions without explicit values, as predicted by more comparative constructive (e.g., MDFT) or heuristic (e.g., Tversky, 1972) theories.

With respect to *action selection*, we find gaze-dependent evidence accumulation as described by the aDDM and the GLAM to accurately capture choices, RTs, and associations with eye tracking measures on the individual level. Concerning the role of *visual attention* in action selection, all our studies align with prior work positively associating visual attention with choice. In Molter and Mohr (2021), we additionally differentiate between effects of viewing duration and order and demonstrate a causal effect of order on action selection. A further aspect of the role of visual attention in the decision process was illustrated in Molter et al. (2021), where gaze was modulated by the context of available alternatives, allowing gaze-dependent evidence accumulation to also account for context-dependence in action selection.

7.4.1 Methodological contribution

In addition to the empirical and theoretical contributions outlined above, the work contained in this thesis also makes a methodological contribution to the field of decision research: The GLAM and the associated open-source Python toolbox (Box 1) enable other researchers to apply the framework of gaze-dependent evidence accumulation in their research. I happily note that these tools are already used by different research groups, addressing a variety of research questions regarding value-based choice, including decision making from memory (Weilbächer et al., 2021), effects of task framing (Sepulveda et al., 2020), mechanisms underlying decision confidence (Brus et al., 2021; Kaanders et al., 2021), and gaze-bias effects in non-human primates (Lupkin & McGinty, 2021).

7.5 Conclusion

To conclude, this thesis offers a refined view on the role of visual attention in constructive preferential choice. Our studies have advanced understanding, raised additional questions, and provided the research community with new tools to address them.

8 Bibliography

Allais, M. (1953). L'Extension des Theories de l'Equilibre Economique General et du Rendement Social au Cas du Risque. *Econometrica*, *21*(2), 269–290. https://doi.org/10.2307/1905539

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*(25), 10367–10371. https://doi.org/10.1073/pnas.1104047108

Anderson, B. (2011). There is no Such Thing as Attention. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00246

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, *3*(5), 8.

Ashby, N. J. S., Jekel, M., Dickert, S., & Glöckner, A. (2016). Finding the right fit: A comparison of process assumptions underlying popular drift-diffusion models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(12), 1982–1993. https://doi.org/10.1037/xlm0000279

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences*, *107*(50), 21767–21772. https://doi.org/10.1073/pnas.0908104107

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*(3), 1331–1348. https://doi.org/10.1037/a0035159

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, *22*(1), 23. https://doi.org/10.2307/1909829

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, *120*(3), 522–543. https://doi.org/10.1037/a0032457

Blakemore, S.-J., & Robbins, T. W. (2012). Decision-making in the adolescent brain. *Nature Neuroscience*, *15*(9), 1184–1191. https://doi.org/10.1038/nn.3177

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*(2), 409–432. https://doi.org/10.1037/0033-295X.113.2.409

Brus, J., Aebersold, H., Grueschow, M., & Polania, R. (2021). Sources of confidence in value-based choice. *PsyArXiv*. https://doi.org/10.31234/osf.io/3wnf7

Busemeyer, J. R., & Diederich, A. (2014). Chapter 4 - Estimation and Testing of Computational Psychological Models. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics (Second Edition)* (pp. 49–61). Academic Press. https://doi.org/10.1016/b978-0-12-416008-8.00004-8

Busemeyer, J. R., Gluth, S., Rieskamp, J., & Turner, B. M. (2019). Cognitive and Neural Bases of Multi-Attribute, Multi-Alternative, Value-based Decisions. *Trends in Cognitive Sciences*, *23*(3), 251–263. https://doi.org/10.1016/j.tics.2018.12.003

Busemeyer, J. R., Johnson, J. G., & Jessup, R. K. (2006). Preferences constructed from dynamic microprocessing mechanisms. In S. Lichtenstein & P. Slovic (Eds.), *The construction of preference*. Cambridge University Press.

Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLOS Computational Biology*, *17*(3), e1008863. https://doi.org/10.1371/journal.pcbi.1008863

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2021). Bambi: A simple interface for fitting Bayesian linear models in Python. *arXiv:2012.10754 [stat]*.

Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology. General*, *143*(4), 1476–1488. https://doi.org/10.1037/a0035813

Chandon, P., Hutchinson, J. W., Bradlow, E. T., & Young, S. H. (2009). Does in-store marketing work? effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *Journal of Marketing*, *73*(6), 1–17. https://doi.org/10.1509/jmkg.73.6.1

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2010). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*(1), 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Ciranka, S., & van den Bos, W. (2021). Adolescent risk-taking in the context of exploration and social influence. *Developmental Review*, *61*, 100979. https://doi.org/10.1016/j.dr.2021.100979

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105–110. https://doi.org/10.1038/nn.3279

Farrell, S., & Lewandowsky, S. (2015). An Introduction to Cognitive Modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An Introduction to Model-Based Cognitive Neuroscience* (pp. 3–24). Springer. https://doi.org/10.1007/978-1-4939-2236-9_1

Farrell, S., & Lewandowsky, S. (2018). *Computational Modeling of Cognition and Behavior:* (First). Cambridge University Press. https://doi.org/10.1017/cbo9781316272503

Fiedler, K. (2000). Beware of samples! a cognitive-ecological sampling approach to judgment biases. *Psychological Review*, *107*(4), 659–676. https://doi.org/10.1037/0033-295x.107.4.659

Fiedler, S., & Glöckner, A. (2012). The Dynamics of Decision Making in Risky Choice: An Eye-Tracking Analysis. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00335

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106–e6115. https://doi.org/10.1073/pnas.1711978115

Fisher, G. (2017). An attentional drift diffusion model over binary-attribute choice. *Cognition*, *168*, 34–45. https://doi.org/10.1016/j.cognition.2017.06.007

Folke, T., Jacobsen, C., Fleming, S. M., & Martino, B. D. (2016). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, *1*(1), 1–8. https://doi.org/10.1038/s41562-016-0002

Friedman, M. (1953). *Essays in positive economics*. University of Chicago Press.

Ghaffari, M., & Fiedler, S. (2018). The Power of Attention: Using Eye Gaze to Predict Other-Regarding and Moral Choices. *Psychological Science*, *29*(11), 1878–1889. https://doi.org/10.1177/0956797618799301

Gigerenzer, G. (2020). How to explain behavior? *Topics in Cognitive Science*, *12*(4), 1363–1381. https://doi.org/10.1111/tops.12480

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality: The adaptive toolbox*. MIT Press.

Glickman, M., Sharoni, O., Levy, D. J., Niebur, E., Stuphorn, V., & Usher, M. (2019). The formation of preference in risky choice. *PLOS Computational Biology*, *15*(8), e1007201. https://doi.org/10.1371/journal.pcbi.1007201

Glimcher, P. W., & Fehr, E. (Eds.). (2014). *Neuroeconomics: Decision making and the brain* (Second edition). Elsevier/AP, Academic Press is an imprint of Elsevier.

Glöckner, A., & Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*(1), 71–98. https://doi.org/10.1002/bdm.684

Gluth, S., Kern, N., Kortmann, M., & Vitali, C. L. (2020). Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nature Human Behaviour*, 1–12. https://doi.org/10.1038/s41562-020-0822-0

Gluth, S., Sommer, T., Rieskamp, J., & Büchel, C. (2015). Effective connectivity between hippocampus and ventromedial prefrontal cortex controls preferential choices from memory. *Neuron*, *86*(4), 1078–1090. https://doi.org/10.1016/j.neuron.2015.04.023

Gluth, S., Spektor, M. S., & Rieskamp, J. (2018). Value-based attentional capture affects multi-alternative decision making. *eLife*, *7*, e39659. https://doi.org/10.7554/eLife.39659

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Gwinn, R., Leber, A. B., & Krajbich, I. (2019). The spillover effects of attentional learning on value-based choice. *Cognition*, *182*, 294–306. https://doi.org/10.1016/j.cognition.2018.10.012

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del R'ıo, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hayden, B. Y., & Niv, Y. (2021). The case against economic values in the orbitofrontal cortex (or anywhere else in the brain). *Behavioral Neuroscience*, *135*(2), 192–201. https://doi.org/10.1037/bne0000448

Hayes, K. J. (1953). The backward curve: A method for the study of learning. *Psychological Review*, *60*(4), 269–275. https://doi.org/10.1037/h0056308

Heathcote, A., Brown, S. D., & Wagenmakers, E.-J. (2015). An Introduction to Good Practices in Cognitive Modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An Introduction to Model-Based Cognitive Neuroscience* (pp. 25–48). Springer. https://doi.org/10.1007/978-1-4939-2236-9_2

Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience*, *9*(6), 467–479. https://doi.org/10.1038/nrn2374

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from Experience and the Effect of Rare Events in Risky Choice. *Psychological Science*, *15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Holmqvist, K. (Ed.). (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, & Psychophysics*, *81*(7), 2288–2303. https://doi.org/10.3758/s13414-019-01846-w

Huber, J., Payne, J. W., & Puto, C. (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research*, *9*(1), 90–98. https://doi.org/10.1086/208899

Hunt, L. T., Malalasekera, W. M. N., de Berker, A. O., Miranda, B., Farmer, S. F., Behrens, T. E. J., & Kennerley, S. W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. *Nature Neuroscience*, *21*(10), 1471–1481. https://doi.org/10.1038/s41593-018-0239-5

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/mcse.2007.55

Huys, Q. J. M., Browning, M., Paulus, M. P., & Frank, M. J. (2021). Advances in the computational understanding of mental illness. *Neuropsychopharmacology*, *46*(1), 3–19. https://doi.org/10.1038/s41386-020-0746-4

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, *19*(3), 404–413. https://doi.org/10.1038/nn.4238

Isham, E. A., & Geng, J. J. (2013). Looking Time Predicts Choice but Not Aesthetic Value. *Plos One*, *8*(8), e71698. https://doi.org/10.1371/journal.pone.0071698

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10), 1489–1506. https://doi.org/10.1016/s0042-6989(99)00163-7

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. https://doi.org/10.1038/35058500

Jang, A. I., Sharma, R., & Drugowitsch, J. (2020). Optimal policy for attention-modulated decisions explains human fixation behavior. *bioRxiv*, 2020.08.04.237057. https://doi.org/10.1101/2020.08.04.237057

Johnson, E. J., & Ratcliff, R. (2014, January 1). Computational and process models of decision making in psychology and behavioral economics. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics (second edition)* (pp. 35–47). Academic Press. https://doi.org/10.1016/B978-0-12-416008-8.00003-6

Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. C. (2008). Process models deserve process data: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, *115*(1), 263–272. https://doi.org/10.1037/0033-295x.115.1.263

Johnson, J. G., & Frame, M. E. (2019). Using process tracing data to define and test process models. *A handbook of process tracing methods* (pp. 374–387). Routledge.

Johnson, M. M. S. (1990). Age differences in decision making: A process methodology for examining strategic information processing. *Journal of Gerontology*, *45*(2), 75–78. https://doi.org/10.1093/geronj/45.2.P75

Kaanders, P., Sepulveda, P., Folke, T., Ortoleva, P., & De Martino, B. (2021). Cherry-picking information: Humans actively sample evidence to support prior beliefs. *bioRxiv*. https://doi.org/10.1101/2021.06.29.450332

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown; Company.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kim, B., Seligman, D., & Kable, J. W. (2012). Preference reversals in decision making under risk are accompanied by changes in attention to different attributes. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00109

Krajbich, I. (2019). Accounting for attention in sequential sampling models of decision making. *Current Opinion in Psychology*, *29*, 6–11. https://doi.org/10.1016/j.copsyc.2018.10.008

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298. https://doi.org/10.1038/nn.2635

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, *3*, 193. https://doi.org/10.3389/fpsyg.2012.00193

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13852–13857. https://doi.org/10.1073/pnas.1101328108

Krauzlis, R. J., Wang, L., Yu, G., & Katz, L. N. (2021). What is attention? *WIREs Cognitive Science*, *n/a*(n/a), e1570. https://doi.org/10.1002/wcs.1570

Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.

Kwak, Y., Payne, J. W., Cohen, A. L., & Huettel, S. A. (2015). The rational adolescent: Strategic information processing during decision making revealed by eye tracking. *Cognitive development*, *36*, 20–30.

Lee, M. D., Criss, A. H., Devezer, B., Donkin, C., Etz, A., Leite, F. P., Matzke, D., Rouder, J. N., Trueblood, J. S., White, C. N., & Vandekerckhove, J. (2019). Robust Modeling in Cognitive Science. *Computational Brain & Behavior*, *2*(3), 141–153. https://doi.org/10.1007/s42113-019-00029-y

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. Sage Publications.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*(1), 46–55. https://doi.org/10.1037/h0031207

Lichtenstein, S., & Slovic, P. (1973). Response-induced reversals of preference in gambling: An extended replication in las vegas. *Journal of Experimental Psychology*, *101*(1), 16–20. https://doi.org/10.1037/h0035472

Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. Cambridge University Press.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*. https://doi.org/10.1017/s0140525x1900061x

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*(37), 13214–13223. https://doi.org/10.1523/jneurosci.1246-11.2011

Liu, H.-Z., Lyu, X.-K., Wei, Z.-H., Mo, W.-L., Luo, J.-R., & Su, X.-Y. (2020). Exploiting the dynamics of eye gaze to bias intertemporal choice. *Journal of Behavioral Decision Making*, *n/a*(n/a). https://doi.org/10.1002/bdm.2219

Liu, H.-Z., Zhou, Y.-B., Wei, Z.-H., & Jiang, C.-M. (2020). The power of last fixation: Biasing simple choices by gaze-contingent manipulation. *Acta Psychologica*, *208*, 103106. https://doi.org/10.1016/j.actpsy.2020.103106

Liversedge, S. P., Gilchrist, I. D., & Everling, S. (Eds.). (2011). *The Oxford handbook of eye movements*. Oxford University Press.

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*(15), 6139–6144. https://doi.org/10.1073/pnas.1217854110

Luce, R. D. (1959). *Individual choice behavior*. John Wiley.

Lupkin, S. M., & McGinty, V. B. (2021). Monkey see, monkey choose: A nonhuman primate model of gaze biases in economic choice. *Poster presented at the Society for Neuroscience Global Connectome Virtual Meeting*.

Mata, R., Josef, A. K., Samanez-Larkin, G. R., & Hertwig, R. (2011). Age differences in risky choice: A meta-analysis. *Annals of the New York Academy of Sciences*, *1235*, 18–29. https://doi.org/10.1111/j.1749-6632.2011.06200.x

Mather, M. (2006). A review of decision-making processes: Weighing the risks and benefits of aging. *When i'm 64* (pp. 145–173). The National Academies Press.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

McFadden, D. (2001). Economic choices. *American Economic Review*, *91*(3), 351–378. https://doi.org/10.1257/aer.91.3.351

McGinty, V. B., Rangel, A., & Newsome, W. T. (2016). Orbitofrontal Cortex Value Signals Depend on Fixation Location during Free Viewing. *Neuron*, *90*(6), 1299–1311. https://doi.org/10.1016/j.neuron.2016.04.045

McKinney, W. (2012). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. "O'Reilly Media, Inc.".

Milosavljevic, M., Navalpakkam, V., Koch, C., & Rangel, A. (2012). Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology*, *22*(1), 67–74. https://doi.org/10.1016/j.jcps.2011.10.002

Molenaar, P. C. M. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement: Interdisciplinary Research and Perspectives*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molter, F., & Mohr, P. N. C. (2021a). Investigating the causal effects of presentation duration, sequence, and format on binary risky choice. https://osf.io/bth3m

Molter, F., & Mohr, P. N. C. (2021b). Presentation order but not duration affects binary risky choice. *PsyArXiv*. https://doi.org/10.31234/osf.io/gcthj

Molter, F., Thomas, A. W., Heekeren, H. R., & Mohr, P. N. C. (2019). GLAMbox: A Python toolbox for investigating the association between gaze allocation and decision behaviour. *PLoS ONE*, *14*(12). https://doi.org/10.1371/journal.pone.0226428

Molter, F., Thomas, A. W., Huettel, S. A., Heekeren, H. R., & Mohr, P. N. C. (2021). Gaze-dependent evidence accumulation predicts multi-alternative risky choice behaviour. *PsyArXiv*. https://doi.org/10.31234/osf.io/x6nbf

Mullett, T. L., & Stewart, N. (2016). Implications of visual attention phenomena for models of preferential choice. *Decision*, *3*(4), 231. https://doi.org/10.1037/dec0000049

Nesselroade, J. R., & Ram, N. (2004). Studying Intraindividual Variability: What We Have Learned That Will Help Us Understand Lives in Context. *Research in Human Development*, *1*(1-2), 9–29. https://doi.org/10.1080/15427609.2004.9683328

Newell, B. R., & Le Pelley, M. E. (2018). Perceptual but not complex moral judgments can be biased by exploiting the dynamics of eye-gaze. *Journal of Experimental Psychology: General*, *147*(3), 409–417. https://doi.org/10.1037/xge0000386

Noguchi, T., & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition*, *132*(1), 44–56. https://doi.org/10.1016/j.cognition.2014.03.006

Noguchi, T., & Stewart, N. (2018). Multialternative decision by sampling: A model of decision making constrained by process data. *Psychological Review*, *125*(4), 512–544. https://doi.org/10.1037/rev0000102

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, *50*(4), 1645–1656. https://doi.org/10.3758/s13428-017-0998-z

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003

Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O., & Hertwig, R. (2018). Prospect theory reflects selective allocation of attention. *Journal of Experimental Psychology: General*, *147*(2), 147–169. https://doi.org/10.1037/xge0000406

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences*, *21*(6), 425–433. https://doi.org/10.1016/j.tics.2017.03.011

Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., & Richardson, D. C. (2015). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(13), 4170–4175. https://doi.org/10.1073/pnas.1415250112

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552. https://doi.org/10.1037/0278-7393.14.3.534

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral Decision Research: A Constructive Processing Perspective. *Annual Review of Psychology*, *43*(1), 87–131. https://doi.org/10.1146/annurev.ps.43.020192.000511

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*(1), 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525

Pisauro, M. A., Fouragnan, E., Retzler, C., & Philiastides, M. G. (2017). Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nature Communications*, *8*(1), 15808. https://doi.org/10.1038/ncomms15808

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556. https://doi.org/10.1038/nrn2357

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295x.85.2.59

Ratcliff, R., & Childers, R. (2015). Individual Differences and Fitting Methods for the Two-Choice Diffusion Model of Decision Making. *Decision (Washington, D.C.)*, *2015*. https://doi.org/10.1037/dec0000030

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Rayner, K. (2009). The 35th sir frederick bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1446–1465. https://doi.org/10.1037/a0013646

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*(3), 631–661. https://doi.org/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*(2), 370–392. https://doi.org/10.1037/0033-295x.108.2.370

Russo, J. E. (2019, June 10). Eye fixations as a process trace. In M. Schulte-Mecklenbeck, A. Kuehberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods : 2nd edition*. Routledge. https://doi.org/10.4324/9781315160559

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, *2*, e55. https://doi.org/10.7717/peerj-cs.55

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, *5*(17), 61–71. https://doi.org/10.2307/2548836

Scheibehenne, B., Rieskamp, J., & Wagenmakers, E.-J. (2013). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological Review*, *120*(1), 39–64. https://doi.org/10.1037/a0030777

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Sepulveda, P., Usher, M., Davies, N., Benson, A. A., Ortoleva, P., & De Martino, B. (2020). Visual attention modulates the integration of goal-relevant evidence and not value. *eLife*, *9*, e60705. https://doi.org/10.7554/eLife.60705

Shadlen, M. N., & Newsome, W. T. (2001). Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey. *Journal of Neurophysiology*, *86*(4), 1916–1936. https://doi.org/10.1152/jn.2001.86.4.1916

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322. https://doi.org/10.1038/nn1150

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, *69*(1), 99–118.

Simonson, I. (1989). Choice Based on Reasons: The Case of Attraction and Compromise Effects. *Journal of Consumer Research*, *16*(2), 158–174. https://doi.org/10.1086/209205

Slovic, P. (1995). The construction of preference. *American psychologist*, *50*(5), 364.

Smith, S. M., & Krajbich, I. (2018). Attention and choice across domains. *Journal of Experimental Psychology: General*, *147*(12), 1810–1826. https://doi.org/10.1037/xge0000482

Smith, S. M., Krajbich, I., & Webb, R. (2019). Estimating the dynamic role of attention via random utility. *Journal of the Economic Science Association*, *5*(1), 97–111. https://doi.org/10.1007/s40881-019-00062-4

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*(7). https://doi.org/10.1371/journal.pcbi.1002607

Stewart, N., Hermens, F., & Matthews, W. J. (2016). Eye Movements in Risky Choice: Eye Movements in Risky Choice. *Journal of Behavioral Decision Making*, *29*(2-3), 116–136. https://doi.org/10.1002/bdm.1854

Sui, X.-Y., Liu, H.-Z., & Rao, L.-L. (2020). The timing of gaze-contingent decision prompts influences risky choice. *Cognition*, *195*, 104077. https://doi.org/10.1016/j.cognition.2019.104077

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second edition). The MIT Press.

Tavares, G., Perona, P., & Rangel, A. (2017). The Attentional Drift Diffusion Model of Simple Perceptual Decision-Making. *Frontiers in Neuroscience*, *11*, 468. https://doi.org/10.3389/fnins.2017.00468

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness* (Rev. and expanded ed). Penguin Books.

Thomas, A. W., Molter, F., & Krajbich, I. (2021). Uncovering the computational mechanisms underlying many-alternative choice. *eLife*, *10*, e57012. https://doi.org/10.7554/eLife.57012

Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R., & Mohr, P. N. C. (2019). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, *3*(6), 625–635. https://doi.org/10.1038/s41562-019-0584-8

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences*, *110*(40), E3858–e3867. https://doi.org/10.1073/pnas.1304429110

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*(24), 9659–9664. https://doi.org/10.1073/pnas.1119569109

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65–79. https://doi.org/10.1016/j.jmp.2016.01.001

Turner, B. M., Schley, D. R., Muller, C., & Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological Review*, *125*(3), 329–362. https://doi.org/10.1037/rev0000089

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281–299. https://doi.org/10.1037/h0032955

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/bf00122574

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*(10), 1179–1189. https://www.jstor.org/stable/2632953

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. https://doi.org/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*(3), 757–769. https://doi.org/10.1037/0033-295x.111.3.757

Vaidya, A. R., & Fellows, L. K. (2015). Testing necessary regional frontal contributions to value assessment and fixation-based updating. *Nature Communications*, *6*, 10120. https://doi.org/10.1038/ncomms10120

van den Bos, W., & Eppinger, B. (2015). Developing developmental cognitive neuroscience: From agenda setting to hypothesis testing. *Developmental Cognitive Neuroscience*, *17*, 138–144. https://doi.org/10.1016/j.dcn.2015.12.011

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, *40*(1), 61–72. https://doi.org/10.3758/brm.40.1.61

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vickers, D. (1979). *Decision processes in visual perception*. Academic Press.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*(1), 37–58.

Vlaev, I., Chater, N., Stewart, N., & Brown, G. D. A. (2011). Does the brain calculate value? *Trends in Cognitive Sciences*, *15*(11), 546–554. https://doi.org/10.1016/j.tics.2011.09.008

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

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767–775. https://doi.org/10.3758/bf03192967

Wagenmakers, E.-J., Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22. https://doi.org/10.3758/bf03194023

Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, *36*(5), 955–968. https://doi.org/10.1016/S0896-6273(02)01092-9

Waskom, M., & seaborn development team, t. (2020). Mwaskom/seaborn. https://doi.org/10.5281/zenodo.592845

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, *60*(1), 53–85. https://doi.org/10.1146/annurev.psych.60.110707.163633

Weilbächer, R. A., Krajbich, I., Rieskamp, J., & Gluth, S. (2021). The influence of visual attention on memory-based preferential choice. *Cognition*, *215*, 104804. https://doi.org/10.1016/j.cognition.2021.104804

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*, 14. https://doi.org/10.3389/fninf.2013.00014

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

Yechiam, E., & Hochman, G. (2013). Loss-aversion or loss-attention: The impact of losses on cognitive performance. *Cognitive Psychology*, *66*(2), 212–231. https://doi.org/10.1016/j.cogpsych.2012.12.001

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

Zilker, V., & Pachur, T. (2021). Nonlinear probability weighting can reflect attentional biases in sequential sampling. *Psychological Review*. https://doi.org/10.1037/rev0000304

9 Appendix

9.1 Deutsche Zusammenfassung

Wie treffen Menschen einfache Entscheidungen, zum Beispiel die Auswahl eines Frühstücks am Hotelbuffet? Anders, als es normative Theorien zur Entscheidungsfindung voraussetzen, sind unsere Präferenzen oft nicht starr, sondern werden erst zum Zeitpunkt der Entscheidung durch die Interaktion des Entscheiders mit seiner Umgebung konstruiert. Empirische Befunde zeigen, dass die Verteilung visueller Aufmerksamkeit während des Entscheidungsprozesses eng mit den getroffenen Entscheidungen zusammenhängt, wobei ein längerer Blick auf eine Alternative mit einer höheren Wahrscheinlichkeit verbunden ist, diese auszuwählen.

Frühere Arbeiten haben die Prozesse, die solchen einfachen Entscheidungen zugrunde liegen, als Akkumulation von Evidenz über Zeit charakterisiert, wobei die Akkumulationsrate zu jedem Zeitpunkt von der Blickrichtung des Entscheiders abhängt. Danach wird eine Entscheidung getroffen, sobald die Evidenz für eine Alternative einen bestimmten Grenzwert überschreitet. Es ist jedoch in mehrerlei Hinsicht unklar, inwiefern diese Theorie von blickabhängiger Evidenzakkumulation generalisierbar ist. Zum einen ist nicht sichergestellt, dass blickabhängige Evidenzakkumulation das Verhalten *einzelner* Entscheider erfasst, und in welchem Ausmaß der Zusammenhang visueller Aufmerksamkeit und Entscheidung zwischen Personen variiert. Zum anderen ist unklar, ob dieser Erklärungsansatz in Situationen Bestand hat, in denen Entscheidungen deutlich von normativen Vorhersagen abweichen. Zuletzt bleibt umstritten, ob visuelle Aufmerksamkeit kausalen Einfluss auf Entscheidungsprozesse hat oder vielmehr die Konstruktion von Präferenzen nur abbildet.

Die vorliegende Dissertation soll diese Fragen in drei empirischen Studien unter Nutzung computerbasierter Modelle des Entscheidungsprozesses beantworten.

In Studie 1 (Molter et al., 2019; Thomas et al., 2019) wurde zunächst ein neuartiges blickabhängiges Evidenzakkumulationsmodell entwickelt, das die Untersuchung des Entscheidungsprozesses einzelner Entscheider erlaubt. Hierzu wurde zusätzlich eine Python-Software-Toolbox veröffentlicht, die auch anderen Forschungsgruppen eine Anwendung des Modells ermöglicht. In vier verschiedenen Datensätzen konnten wir mithilfe dieses Werkzeugs zeigen, dass blickabhängige Evidenzakkumulation präzise Vorhersagen über Entscheidungen, Antwortzeiten und deren Zusammenhänge mit visueller Aufmerksamkeit für einzelne Entscheider macht. Unsere Analysen zeigten jedoch auch, dass Individuen große Unterschiede beim Zusammenhang von visueller Aufmerksamkeit und Entscheidung aufwiesen. Diese individuellen Unterschiede gingen zudem mit individuellen Unterschieden in der Konsistenz, mit der Entscheidungen getroffen wurden, einher.

In Studie 2 (Molter et al., 2021) wurde das Konzept blickabhängiger Evidenzakkumulation in einer Entscheidungsaufgabe geprüft, in der drei risikobehaftete Lotterien als Alternativen mit mehreren Attributen zur Auswahl standen. Die Aufgabe wurde so entwickelt, dass Kontexteffekte im Entscheidungsverhalten auftreten sollten. Kontexteffekte beschreiben Präferenzänderungen in Abhängigkeit der verfügbaren Alternativen und stellen starke Abweichungen von normativen Vorhersagen dar. Die Ergebnisse zeigten, dass nicht nur das Entscheidungsverhalten, sondern auch die Verteilung visueller Aufmerksamkeit vom Kontext der verfügbaren Alternativen moduliert wurde. Dies ermöglichte es einem aus Vorarbeiten abgeleiteten blickabhängigen Evidenzakkumulationsmodell, Entscheidungsverhalten in diesem komplexen Szenario zu erfassen.

Zuletzt wurde in einer prä-registrierten dritten Studie (Molter & Mohr, 2021b) die Richtung des Kausalitätszusammenhangs zwischen visueller Aufmerksamkeit und Entscheidung beleuchtet. In unserem Experiment trafen Teilnehmer wiederholte Entscheidungen zwischen zwei Lotterien, deren Attribute sequenziell präsentiert wurden. Dies ermöglichte die experimentelle Kontrolle von Präsentationsdauer und Reihenfolge der Stimulusinformation. Die Ergebnisse bestätigten einen kausalen

Einfluss der Informationssuche auf die Präferenzkonstruktion. Jedoch wurde hier die Präsentationsreihenfolge, nicht die Präsentationsdauer als Einflussfaktor identifiziert. Bemerkenswert ist hierbei, dass nur manche blickabhängigen Evidenzakkumulationsmodelle solche kausalen Einflüsse der Reihenfolge vorhersagen. Unsere Ergebnisse zeigen dementsprechend ein mögliches Potenzial für zukünftige Theorieentwicklung auf.

Unsere Studien bestätigten grundsätzlich den positiven Zusammenhang zwischen visueller Aufmerksamkeit und Entscheidungen. Zudem unterstützen sie Theorien blickabhängiger Evidenzakkumulation im Rahmen individueller und komplexer Entscheidungen. Die Analysen haben allerdings auch bedeutende individuelle Unterschiede und mögliche Grenzen aktueller Modelle sichtbar gemacht. Hier konnten wir jedoch zeigen, dass die Berücksichtigung solcher Unterschiede und die Hinzunahme zusätzlicher Mechanismen wie imperfekter Akkumulation die Vorhersage individuellen Verhaltens erheblich verbessert.

Zum Abschluss der Arbeit werden diese Ergebnisse einer aktiven Rolle visueller Aufmerksamkeit im Entscheidungsprozess sowie das theoretische Modell blickabhängiger Evidenzakkumulation im weiteren Kontext konstruierter Präferenzen diskutiert und mögliche Implikationen für die computermodellbasierte Analyse von Entscheidungs- und Blickbewegungsdaten aufgezeigt.

9.2 List of publications

1. Molter, F., & Mohr, P. N. C. (2021b). Presentation order but not duration affects binary risky choice. *PsyArXiv*. https://doi.org/10.31234/osf.io/gcthj

2. Thomas, A. W., Molter, F., & Krajbich, I. (2021). Uncovering the computational mechanisms underlying many-alternative choice. *eLife*, *10*, e57012. https://doi.org/10.7554/eLife.57012

3. Molter, F., Thomas, A. W., Huettel, S. A., Heekeren, H. R., & Mohr, P. N. C. (2021). Gaze-dependent evidence accumulation predicts multi-alternative risky choice behaviour. *PsyArXiv*. https://doi.org/10.31234/osf.io/x6nbf

4. Molter, F., Thomas, A. W., Heekeren, H. R., & Mohr, P. N. C. (2019). GLAMbox: A Python toolbox for investigating the association between gaze allocation and decision behaviour. *PLoS ONE*, *14*(12). https://doi.org/10.1371/journal.pone.0226428

5. Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R., & Mohr, P. N. C. (2019). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, *3*(6), 625–635. https://doi.org/10.1038/s41562-019-0584-8

6. Bellucci, G., Molter, F., & Park, S. Q. (2019). Neural representations of honesty predict future trust behavior. *Nature Communications*, *10*(1), 1–12. https://doi.org/10.1038/s41467-019-13261-8

9.3 Talks and presentations

October 2017 Toronto, Canada, 15th Annual Meeting of the Society of Neuroeconomics. Accounting for individual differences in gaze-weighted evidence accumulation improves prediction of individual consumer choice (Poster presentation).

October 2017 Toronto, Canada, *Rotman School of Management*, 6th Consumer Neuroscience Satellite Symposium of the Annual Conference of the Society for Neuroeconomics. Accounting for individual differences in gaze-weighted evidence accumulation improves prediction of individual consumer choice.

November 2016 Boston, Massachusetts, USA, Society for Judgment and Decision Making. The 2016 37th Annual Conference. h-aDDM: A hierarchical framework to model economic choices and eye movements (Poster presentation).

October 2016 Berlin, Germany, International Max Planck Research School LIFE Academy. The h-aDDM toolbox.

August 2016 Berlin, Germany, *WZB Berlin Social Science Center*, Consumer Neuroscience Satellite Symposium of the Annual Conference of the Society for Neuroeconomics. The role of visual attention in contextual risky choice.

June 2016 Bonn, Germany, *Max Planck Institute for Research on Collective Goods*, 35th Meeting of the European Group of Process Tracing Studies (EG-PROC). The role of visual attention in contextual risky choice.

May 2016 Charlottesville, Virginia, USA, IMPRS-LIFE Academy. Can attention-based decision models explain developmental differences in risky choice?

November 2015 Berlin, Germany, *WZB Berlin Social Science Center*, Interdisciplinary Perspectives on Decision Making. The role of visual attention in contextual risky choice.

September 2015 **Miami, Florida, USA,** 13th Annual Meeting of the Society for Neuroeconomics. The role of eye movements in contextual risky choice (Poster presentation).

May 2015 **Ann Arbor, Michigan, USA,** International Max Planck Research School LIFE Academy. The role of eye movements in contextual risky choice (Poster presentation).

9.4 Eigenanteil

Erklärung gemäß § 7 Abs. 3 Satz 4 der Promotionsordnung über den Eigenanteil an den veröffentlichten oder zur Veröffentlichung vorgesehenen eingereichten wissenschaftlichen Schriften im Rahmen meiner publikationsbasierten Arbeit

I. Name, Vorname: Molter, Felix
Institut: Fachbereich Erziehungswissenschaft und Psychologie
Promotionsfach: Psychologie
Titel: Master of Science (Social, Cognitive and Affective Neuroscience)

II. Nummerierte Aufstellung der eingereichten Schriften (Titel, Autoren, wo und wann veröffentlicht bzw. eingereicht):

1. Thomas, A. W.[1], **Molter, F.**[1], Krajbich, I., Heekeren, H. R., & Mohr, P. N. C. (2019). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, 3(6), 625-635. doi.org/10.1038/s41562-019-0584-8
[1]Shared first authorship with equal contribution.

2. **Molter, F.**[1], Thomas, A. W.[1], Heekeren, H. R., & Mohr, P. N. C. (2019). GLAMbox: A Python toolbox for investigating the association between gaze allocation and decision behaviour. *PLOS ONE*, 14(12), e0226428. doi.org/10.1371/journal.pone.0226428
[1]Shared first authorship with equal contribution.

3. **Molter, F.**, Thomas, A., Huettel, S. A., Heekeren, H. R., & Mohr, P. N. C. (submitted). Gaze-dependent evidence accumulation predicts multi-alternative risky choice behaviour. doi.org/10.31234/osf.io/x6nbf

4. **Molter, F.** & Mohr, P. N. C. (submitted). Presentation order but not duration affects binary risky choice. doi.org/10.31234/osf.io/gcthj

III. Darlegung des eigenen Anteils an diesen Schriften:

Die Bewertung des Eigenanteils erfolgt auf der Skala: "führend – maßgeblich – geringfügig".

Zu II. 1.: Konzeption (maßgeblich), Methodenentwicklung (maßgeblich), Datenauswertung (maßgeblich), Programmierung (maßgeblich), maßgeblich (mehrheitlich), Erstellen des Manuskriptes (maßgeblich).

Zu II. 2.: Konzeption (maßgeblich), Methodenentwicklung (maßgeblich), Datenauswertung (maßgeblich), Programmierung & Dokumentation (maßgeblich), Ergebnisdiskussion (maßgeblich), Erstellen des Manuskriptes (maßgeblich).

Zu II. 3.: Konzeption (führend), Versuchsdesign (führend), Datenerhebung (führend), Methodenentwicklung (führend), Datenauswertung (führend), Programmierung (führend), Ergebnisdiskussion (maßgeblich), Erstellen des Manuskriptes (führend).

Zu II. 4.: Konzeption (führend), Versuchsdesign (führend), Datenerhebung (führend), Methodenentwicklung (führend), Datenauswertung (führend), Programmierung (führend), Ergebnisdiskussion (maßgeblich), Erstellen des Manuskriptes (führend).

9.5 Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt,

- dass ich die vorliegende Arbeit selbstständig und ohne unerlaubte Hilfe verfasst habe,
- dass ich mich nicht bereits anderwärts um einen Doktorgrad beworben habe und keinen Doktorgrad in dem Promotionsfach Psychologie besitze und
- dass ich die zugrunde liegende Promotionsordnung vom 08. August 2016 kenne.

Berlin, den 30. September 2021

Felix Molter

9.6 Research articles

Thomas, A. W.[1], **Molter, F.**[1], Krajbich, I., Heekeren, H. R., & Mohr, P. N. C. (2019). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, 3(6), 625-635. https://doi.org/10.1038/s41562-019-0584-8

[1]Shared first authorship with equal contribution.

Molter, F.[1], Thomas, A. W.[1], Heekeren, H. R., & Mohr, P. N. C. (2019). GLAMbox: A Python toolbox for investigating the association between gaze allocation and decision behaviour. *PLOS ONE*, 14(12), e0226428. https://doi.org/10.1371/journal.pone.0226428

[1]Shared first authorship with equal contribution.

PLOS ONE

RESEARCH ARTICLE

GLAMbox: A Python toolbox for investigating the association between gaze allocation and decision behaviour

Felix Molter[1,2,3,4☯]*, **Armin W. Thomas**[2,3,5,6☯]*, **Hauke R. Heekeren**[2,3], **Peter N. C. Mohr**[1,2,4]

1 WZB Berlin Social Science Center, Berlin, Germany, **2** Center for Cognitive Neuroscience Berlin, Freie Universität Berlin, Berlin, Germany, **3** Department of Education and Psychology, Freie Universität Berlin, Germany, **4** School of Business and Economics, Freie Universität Berlin, Germany, **5** Department of Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany, **6** Max Planck School of Cognition, Leipzig, Germany

☯ These authors contributed equally to this work.
* felixmolter@gmail.com (FM); athms.research@gmail.com (AT)

Check for updates

OPEN ACCESS

Citation: Molter F, Thomas AW, Heekeren HR, Mohr PNC (2019) GLAMbox: A Python toolbox for investigating the association between gaze allocation and decision behaviour. PLoS ONE 14(12): e0226428. https://doi.org/10.1371/journal.pone.0226428

Editor: David V. Smith, Temple University, UNITED STATES

Received: August 26, 2019

Accepted: November 26, 2019

Published: December 16, 2019

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0226428

Copyright: © 2019 Molter et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All files are available at Github under https://github.com/glamlab/glambox.

Abstract

Recent empirical findings have indicated that gaze allocation plays a crucial role in simple decision behaviour. Many of these findings point towards an influence of gaze allocation onto the speed of evidence accumulation in an accumulation-to-bound decision process (resulting in generally higher choice probabilities for items that have been looked at longer). Further, researchers have shown that the strength of the association between gaze and choice behaviour is highly variable between individuals, encouraging future work to study this association on the individual level. However, few decision models exist that enable a straightforward characterization of the gaze-choice association at the individual level, due to the high cost of developing and implementing them. The model space is particularly scarce for choice sets with more than two choice alternatives. Here, we present GLAMbox, a Python-based toolbox that is built upon PyMC3 and allows the easy application of the gaze-weighted linear accumulator model (GLAM) to experimental choice data. The GLAM assumes gaze-dependent evidence accumulation in a linear stochastic race that extends to decision scenarios with many choice alternatives. GLAMbox enables Bayesian parameter estimation of the GLAM for individual, pooled or hierarchical models, provides an easy-to-use interface to predict choice behaviour and visualize choice data, and benefits from all of PyMC3's Bayesian statistical modeling functionality. Further documentation, resources and the toolbox itself are available at https://glambox.readthedocs.io.

Introduction

A plethora of empirical findings has established an association between gaze allocation and decision behaviour on the group-level. For example, in value-based decision making, it has been repeatedly shown that longer gaze towards one option is associated with a higher choice

Funding: The Junior Professorship of P.N.C.M. as well as the associated Dahlem International Network Junior Research Group Neuroeconomics is supported by Freie Universität Berlin within the Excellence Initiative of the German Research Foundation (DFG). Further support for P.N.C.M. is provided by the WZB Berlin Social Science Center. F.M. is supported by the International Max Planck Research School on the Life Course (LIFE). A.T. is supported by the BMBF and Max Planck Society. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

probability for that option [1–13] and that external manipulation of gaze allocation changes choice probabilities accordingly [1, 9, 10, 14]. Such gaze bias effects are not limited to value-based decisions, but have recently also been observed in perceptual choices, where participants judge the perceptual attributes of stimuli based on available sensory information [14].

These findings have led to the development of a set of computational models, aimed at capturing the empirically observed association between gaze allocation and choice behaviour by utilizing gaze data to inform the momentary accumulation rates of diffusion decision processes [2, 7, 8, 14–17]. Specifically, these models assume that evidence accumulation in favour of an item continues while it is not looked at, but at a discounted rate. The application of these models is limited so far, as fitting them to empirical data depends on computationally expensive simulations, involving the simulation of fixation trajectories. These simulations, as well as the creation of models of the underlying fixation process, become increasingly difficult with increasing complexity of the decision setting (e.g., growing choice set sizes or number of option attributes, etc). Existing approaches that circumvent the need for simulations, model the evidence accumulation process as a single diffusion process between two decision bounds and are therefore limited to binary decisions [2, 18].

However, researchers are increasingly interested in choice settings involving more than two alternatives. Choices outside the laboratory usually involve larger choice sets or describe items on multiple attributes. Besides, many established behavioural effects only occur in multi-alternative and multi-attribute choice situations [19].

Furthermore, recent findings indicate strong individual differences in the association between gaze allocation and choice behaviour [20, 21] as well as individual differences in the decision mechanisms used [15]. While the nature of individual differences in gaze biases is still not fully understood, different mechanisms have been suggested: Smith and Krajbich [20] showed that gaze bias differences can be related to individual differences in attentional scope ("tunnel vision"). Vaidya and Fellows [13] found stronger gaze biases in patients with damage in dorsomedial prefrontal cortex (PFC). Further, recent empirical work has investigated the roles of learning and attitude accessibility in gaze dependent decision making [22, 23]. However, more systematic investigations of these differences are needed, as the majority of model-based investigations of the relationship between gaze allocation and choice behaviour were focused on the group level, disregarding differences between individuals.

With the Gaze-weighted linear accumulator model (GLAM; [21]), we have proposed an analytical tool that allows the model-based investigation of the relationship between gaze allocation and choice behaviour at the level of the individual, in choice situations involving more than two alternatives, solely requiring participants' choice, response time (RT) and gaze data, in addition to estimates of the items' values.

Like the attentional Drift Diffusion Model (aDDM) [7, 8, 17], the GLAM assumes that the decision process is biased by momentary gaze behaviour: While an item is not fixated, its value representation is discounted. The GLAM, however, differs from the aDDM in other important aspects: In contrast to the aDDM, the fixation-dependent value signals are averaged across the trial, using the relative amount of time individuals spend fixating the items. This step abstracts away the specific sequence of fixations in a trial, that can be investigated with the aDDM. On the other hand, this simplification allows for the construction of trial-wise constant drift rates that can enter a basic stochastic race framework. While race models like the GLAM are not statistically optimal [24] the GLAM has been shown to provide a good fit to empirical data [21]. In general, race models have at least two practical advantages: First, they often have analytical solutions to their first-passage density distributions, and secondly, they naturally generalize to choice scenarios involving more than two alternatives. The analytical tractability of the race framework further allows for efficient parameter estimation in a hierarchical Bayesian manner. The GLAM thereby integrates gaze-dependent accumulation into a practical race model shell.

To make GLAM more accessible, we now introduce GLAMbox, a Python-based toolbox for the application of the GLAM to empirical choice, RT and gaze data. GLAMbox allows for individual and hierarchical estimation of the GLAM parameters, simulation of response data and model-based comparisons between experimental conditions and groups. It further contains a set of visualization functions to inspect choice and gaze data and evaluate model fit. We illustrate three application examples of the toolbox: In Example 1, we illustrate how GLAMbox can be used to analyze individual participant data with the GLAM. In particular, we perform an exemplary model comparison between multiple model variants on the individual level, as well as an out-of-sample prediction of participants' choice and RT data. In Example 2, we demonstrate the application of the GLAM to perform a comparison of group-level parameters in a setting with limited amounts of data, using hierarchical parameter estimation. Lastly, in Example 3, we walk the reader through a step-by-step parameter recovery study with the GLAM, which is encouraged to increase confidence in the estimated parameter values.

Materials and methods

Gaze-weighted linear accumulator model details

Like the aDDM, the GLAM assumes that preference formation, during a simple choice process, is guided by the allocation of visual gaze (for an overview, see Fig 1). Particularly, the decision process is guided by a set of decision signals: An absolute and relative decision signal. Throughout the trial, the absolute signal of an item i can be in two states: An unbiased state,

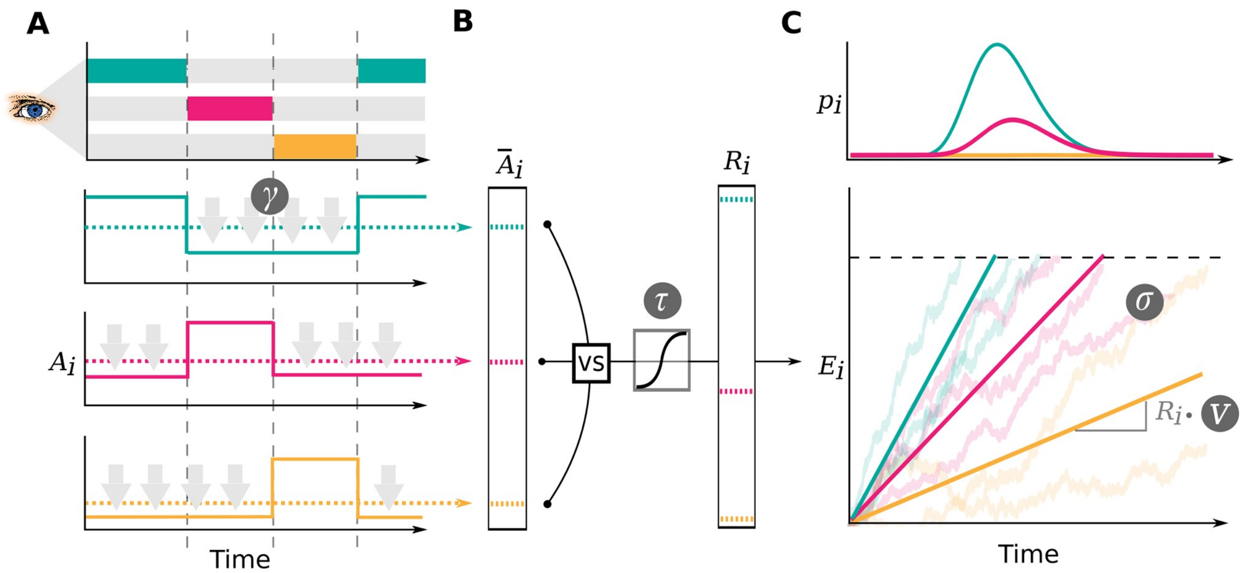


Fig 1. Gaze-weighted linear accumulator model. In the GLAM, preference formation during the decision process is dependent on the allocation of visual gaze (A). For each item in the choice set, an average absolute decision signal $\bar{A}_i$ is computed (dashed lines in A). The magnitude of this signal is determined by the momentary allocation of visual gaze: While an item is currently not looked at, its signal is discounted by parameter γ ($\gamma \leq 1$; discounting is illustrated by gray arrows) (A). To determine a relative decision signal R_i for each item in the choice set, absolute evidence signals are transformed in two steps (B): First, the difference between each average absolute decision signal $\bar{A}_i$ and the maximum of all others is determined. Second, the resulting differences are scaled through a logistic transform, as the GLAM assumes an adaptive representation of the relative decision signals that is especially sensitive to differences close to 0 (where the absolute signal for an item is very close to the maximum of all others). The resulting relative decision signals R_i can be used to predict choice and RT, by determining the speed of the accumulation process in a linear stochastic race (C). The stochastic race then provides first-passage time distributions p_i, describing the likelihood of each item being chosen at each time point.

https://doi.org/10.1371/journal.pone.0226428.g001

equal to the item's value r_i while the item is looked at, and a biased state while any other item is looked at, where the item value r_i is discounted by a parameter γ. The average absolute decision signal $\bar{A}_i$ is given by

$$\bar{A}_i = g_i r_i + (1 - g_i)\gamma r_i, \tag{1}$$

where g_i is defined as the fraction of total trial time that item i was looked at. If $\gamma = 1$, there is no difference between the biased and unbiased state, resulting in no influence of gaze allocation on choice behaviour. For γ values less than 1, the absolute decision signal A_i is discounted, resulting in generally higher choice probabilities for items that have been looked at longer. For γ values less than 0, the sign of the absolute decision signal A_i changes, when the item is not looked at, leading to an overall even stronger gaze bias, as evidence for these items is actively lost, when they are not looked at. This type of gaze-dependent leakage mechanism is supported by a variety of recent empirical findings [15, 21].

To determine the relative decision signals, the average absolute decision signals $\bar{A}_i$ are transformed in two steps: First, for each item i, the relative evidence R_i^* is computed as the difference between the average absolute decision signal of the item $\bar{A}_i$ (Eq 1) and the maximum of all other average absolute decision signals $\bar{A}_{j \neq i}$ (also obtained from Eq 2) is computed:

$$R_i^* = \bar{A}_i - \max_{j \neq i} \bar{A}_j. \tag{2}$$

Second, the resulting difference signals R_i^* are scaled through a logistic transform $s(x)$. The GLAM assumes an adaptive representation of the relative decision signals, which is maximally sensitive to small differences in the absolute decision signals close to 0 (where the difference between the absolute decision signal of an item and the maximum of all others is small):

$$R_i = s(R_i^*) \tag{3}$$

$$s(x) = \frac{1}{1 + \exp(-\tau x)} \tag{4}$$

The sensitivity of this transform is determined by the temperature parameter τ of the logistic function. Larger values of τ indicate stronger sensitivity to small differences in the average absolute decision signals $\bar{A}_i$.

Unlike more traditional diffusion models (including the aDDM), the GLAM employs a linear stochastic race to capture response behaviour as well as RTs. The relative signals R_i enter a race process, where one item accumulator E_i is defined for each item in the choice set:

$$E_i(t) = E_i(t-1) + vR_i + N(0, \sigma^2), \text{with } E_i(0) = 0 \tag{5}$$

At each time step t, the amount of accumulated evidence is determined by the accumulation rate vR_i, and zero-centered normally distributed noise with standard deviation σ. The velocity parameter v linearly scales the item drift rates in the race process and thereby affects the response times produced by the model: Lower values of v produce longer response times, larger vs result in shorter response times. A choice for an item is made as soon as one accumulator reaches the decision boundary b. To avoid underdetermination of the model, either the velocity parameter v, the noise parameter σ or the boundary has to be fixed. Similar to the aDDM, the GLAM fixes the boundary to a value of 1. The first passage time density $f_i(t)$ of a single linear stochastic accumulator E_i, with decision boundary b, is given by the inverse

Gaussian distribution:

$$f_i(t) = \left[\frac{\lambda}{2\pi t^3}\right]^{\frac{1}{2}} \exp\left(\frac{-\lambda(t-\mu)^2}{2\mu^2 t}\right) \tag{6}$$

$$\text{with } \mu = \frac{b}{vR_i} \text{ and } \lambda = \frac{b^2}{\sigma^2}$$

However, this density does not take into account that there are multiple accumulators in each trial racing towards the same boundary. For this reason, $f_i(t)$ must be corrected for the probability that any other accumulator crosses the boundary first. The probability that an accumulator crosses the boundary prior to t, is given by its cumulative distribution function $F_i(t)$:

$$F_i(t) = \Phi\left(\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu}-1\right)\right) + \exp\left(\frac{2\lambda}{\mu}\right)\cdot\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu}+1\right)\right) \tag{7}$$

Here, $\Phi(x)$ defines the standard normal cumulative distribution function. Hence, the joint probability $p_i(t)$ that accumulator E_i crosses b at time t, and that no other accumulator $E_{j\neq i}$ has reached b first, is given by:

$$p_i(t) = f_i(t)\prod_{j\neq i}(1 - F_j(t)) \tag{8}$$

Contaminant response model. To reduce the influence of erroneous responses (e.g., when the participant presses a button by accident or has a lapse of attention during the task) on parameter estimation, we include a model of contaminant response processes in all estimation procedures: In line with existing drift diffusion modelling toolboxes [25], we assume a fixed 5% rate of erroneous responses ϵ that is modeled as a participant-specific uniform likelihood distribution $u_s(t)$. This likelihood describes the probability of a random choice for any of the N available choice items at a random time point in the interval of empirically observed RTs [25, 26]:

$$u_s(t) = \frac{1}{N(\max \mathrm{rt}_s - \min \mathrm{rt}_s)} \tag{9}$$

The resulting likelihood for participant s choosing item i, accounting for erroneous responses, is then given by:

$$l_i(t) = (1-\epsilon)\cdot p_i(t) + \epsilon\cdot u_s(t) \tag{10}$$

The rate of error responses ϵ can be specified by the user to a different value than the default of 5% using the `error_weight` keyword in the `make_model` method (see below).

Individual parameter estimation details. The GLAM is implemented in a Bayesian framework using the Python library PyMC3 [27]. The model has four parameters (v, γ, σ, τ). By default, uninformative, uniform priors between sensible limits (derived from earlier applications to four different datasets: [21]) are placed on all parameters:

$$
\begin{aligned}
v &\sim U(0,4)\\
\gamma &\sim U(-2,1)\\
\sigma &\sim U(0,4)\\
\tau &\sim U(0,10)
\end{aligned}
$$

These limits were derived by extending the range of observed parameter estimates in earlier applications of the GLAM to four different empirical choice datasets. These datasets encompass data of 117 participants in value-based and perceptual choice tasks with up to three choice alternatives (including a wide range of possible response times, gaze bias strengths and choice accuracies; for further details [21]). Parameter estimates for these datasets are illustrated and summarised in S1 Table, S1 Fig and S1 Fig.

The velocity parameter v and the noise parameter σ must be strictly positive. Smaller v produce slower and less accurate responses (for constant σ), while smaller σ produce more accurate and slower responses (for constant v). The gaze bias parameter γ has a natural upper bound at 1 (indicating no gaze bias), while decreasing γ values indicate an increasing gaze bias strength. The sensitivity parameter τ has a natural lower bound at 0 (resulting in no sensitivity to differences in average absolute decision signals $\bar{A}_i$), with larger values indicating increased sensitivity.

Hierarchical parameter estimation details. For hierarchical models, individual parameters are assumed to be drawn from Truncated Normal distributions, parameterized by mean and standard deviation, over which weakly informative, Truncated Normal priors are assumed (based on the distribution of group level parameter estimates obtained from four different datasets in [21]; see Fig 2, S1 and S2 Figs and S1 Table):

$$
\begin{aligned}
v_\mu &\sim N(0.63, 10 \cdot 0.26), \text{truncated to } [0,2]\\
v_\sigma &\sim N(0.26, 10 \cdot 0.11), \text{truncated to } [0,1]\\
\gamma_\mu &\sim N(0.12, 10 \cdot 0.11), \text{truncated to } [-2,1]\\
\gamma_\sigma &\sim N(0.35, 10 \cdot 0.1), \text{truncated to } [0,1]\\
\sigma_\mu &\sim N(0.27, 10 \cdot 0.08), \text{truncated to } [0,1]\\
\sigma_\sigma &\sim N(0.05, 10 \cdot 0.01), \text{truncated to } [0,0.2]\\
\tau_\mu &\sim N(1.03, 10 \cdot 0.58), \text{truncated to } [0,5]\\
\tau_\sigma &\sim N(0.62, 10 \cdot 0.26), \text{truncated to } [0,3]
\end{aligned}
$$

Basic usage

Data format, the `GLAM` class. The core functionality of the GLAMbox is implemented in the `GLAM` model class. To apply the GLAM to data, an instance of the model class needs to be instantiated and supplied with the experimental data, first:

```
import glambox as gb
glam = gb.GLAM(data=data)
```

The data must be a pandas [28] DataFrame with one row per trial, containing the following variable entries:

- `subject`: Subject index (integer, first subject should be 0)

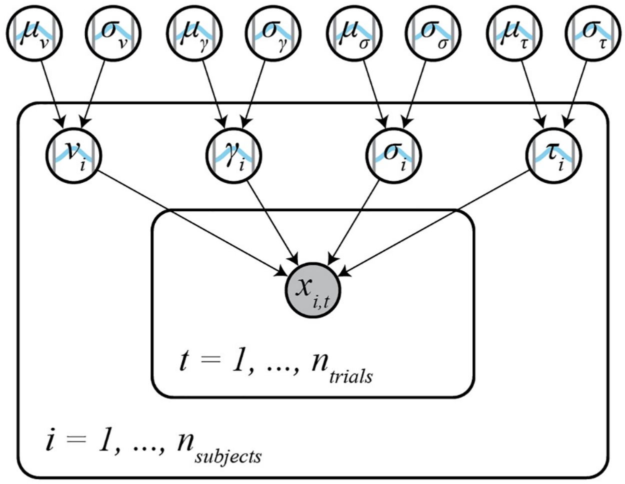


Fig 2. Hierarchical model structure. In the hierarchical model, individual subject parameters γ_i, v_i. σ_i, and τ_i (subject plate) are drawn from Truncated Normal group level distributions with means μ and standard deviations σ (outside of the subject plate). Weakly informative Truncated Normal priors are placed on the group level parameters. RT and choice data $x_{i,t}$ for each trial t is distributed according to the subject parameters and the GLAM likelihood (Eq (8); inner trial plate).

https://doi.org/10.1371/journal.pone.0226428.g002

- `trial`: Trial index (integer, first trial should be 0)
- `choice`: Chosen item in this trial (integer, items should be 0, 1, . . ., N)
- `rt`: Response time (float, in seconds)
- for each item i in the choice set:
  - `item_value_i`: The item value (float, we recommend to re-scale all item values to a range between 1 and 10 to allow comparison of parameter estimates between studies)
  - `gaze_i`: The fraction of total time in this trial that the participant spent looking at this item (float, between 0 and 1)
- additional variables coding groups or conditions (string or integer)

For reference, the first two rows of a pandas DataFrame ready to be used with GLAMbox are shown in Table 1.

Next, the respective PyMC3 model, which will later be used to estimate the model's parameters, can be built using the `make_model` method. Here, the researcher specifies the kind of the model: *`'individual'`* if the parameters should be estimated for each subject individually, *`'hierarchical'`* for hierarchical parameter estimation, or *`'pooled'`* to estimate a single parameter set for all subjects. At this stage, the researcher can also specify experimental parameter dependencies: For example, a parameter could be expected to vary between groups or conditions. In line with existing modeling toolboxes (e.g., [25, 29]) dependencies are

Table 1. The first two rows of a pandas DataFrame ready to be used with GLAM.

subject	trial	choice	rt	item_value_0	item_value_1	item_value_2	gaze_0	gaze_1	gaze_2	speed
0	0	0	2.056	5	1	3	0.16	0.62	0.22	'fast'
0	1	2	3.685	3	6	9	0.44	0.22	0.34	'slow'

https://doi.org/10.1371/journal.pone.0226428.t001

defined using the `depends_on` argument. `depends_on` expects a dictionary with parameters as keys and experimental factors as values (e.g., `depends_on=dict(v='speed')` for factor `'speed'` with conditions `'fast'` and `'slow'` in the data). The toolbox internally handles within- and between subject designs and assigns parameters accordingly. If multiple conditions are given for a factor, one parameter will be designated for each condition. Finally, the `make_model` method allows parameters to be fixed to a specific value using the `*_val` arguments (e.g., `gamma_val=1` for a model without gaze bias). If parameters should be fixed for individual subjects, a list of individual values needs to be passed.

```
model.make_model(kind='individual',
                 depends_on=dict(v='speed'),
                 gamma_val=1)
```

Inference. Once the PyMC3 model is built, parameters can be estimated using the `fit` method:

```
model.fit(method='MCMC')
```

The `fit` method defaults to Markov-Chain-Monte-Carlo (MCMC; [30]) sampling, but also allows for Variational Inference (see below).

Markov-Chain-Monte-Carlo. MCMC methods approximate the Bayesian posterior parameter distributions, describing the probability of a parameter taking certain values given the data and prior probabilities, through repeated sampling. GLAMbox can utilize all available MCMC step methods provided by PyMC3. The resulting MCMC traces can be accessed using the `trace` attribute of the model instance (note that a list of traces is stored for models of kind `'individual'`). They should always be checked for convergence, to ascertain that the posterior distribution is approximated well. Both qualitative visual and more quantitative numerical checks of convergence, such as the Gelman-Rubin statistic $\hat{R}$ and the number of effective samples are recommended (for detailed recommendations, see [31, 32]). PyMC3 contains a range of diagnostic tools to perform such checks (such as the `summary` function).

Variational inference. Estimation can also be done using all other estimation procedures provided in the PyMC3 library. This includes variational methods like Automatic Differentiation Variational Inference (ADVI; [33]). To use variational inference, the `method` argument can be set to `'VI'`, defaulting to the default variational method in PyMC3. We found variational methods to quickly yield usable, but sometimes inaccurate parameter estimates, and therefore recommend using MCMC for final analyses.

Accessing parameter estimates. After parameter estimation is completed, the resulting estimates can be accessed with the `estimates` attribute of the GLAM model instance. This returns a table with one row for each set of parameter estimates for each individual and condition in the data. For each parameter, a *maximum a posteriori* (MAP) estimate is given, in addition to the 95% Highest-Posterior Density Interval (HPD). If the parameters were estimated hierarchically, the table also contains estimates of the group-level parameters.

Comparing parameters between groups or conditions. Parameter estimates can be compared between different experimental groups or conditions (specified with the `depends_on` keyword when calling `make_model`) using the `compare_parameters` function from the `analysis` module. It takes as input the fitted GLAM instance, a list of parameters (`'v'`, `'s'`, `'gamma'`, `'tau'`), and a list of pairwise comparisons between groups or conditions. The comparison argument expects a list of tuples (e.g., `[('group1', 'group2'), ('group1', 'group3')]`). For example, given a fitted model instance (here `glam`) a comparison of the γ parameter between two groups (`group1` and `group2`) can be computed as:

```
from gb.analysis import compare_parameters
comparison = compare_parameters(model=glam,
                                parameters=['gamma'],
                                comparisons=[('group1', 'group2')])
```

The function then returns a table with one row per specified comparison, and columns containing the mean posterior difference, percentage of the posterior above zero, and corresponding 95% HPD interval. If supplied with a hierarchical model, the function computes differences between group-level parameters. If an individual type model is given, it returns comparison statistics for each individual.

Comparisons can be visualized using the `compare_parameters` function from the `plots` module. It takes the same input as its analogue in the `alysis` module. It plots posterior distributions of parameters and the posterior distributions of any differences specified using the `comparisons` argument. For a usage example and plot see Example 2.

Comparing model variants. Model comparisons between multiple GLAM variants (e.g., full and restricted variants) can be performed using the `compare_models` function, which wraps the function of the same name from the PyMC3 library. The `compare_models` function takes as input a list of fitted model instances that are to be compared. Additional keyword arguments can be given and are passed on to the underlying PyMC3 `compare` function. This allows the user, for example, to specify the information criterion used for the comparison via the `ic` argument (`'WAIC'` or `'LOO'` for Leave-One-Out cross validation). It returns a table containing an estimate of the specified information criterion, standard errors, difference to the best-fitting model, standard error of the difference, and other output variables from PyMC3 for each inputted model (and subject, if individually estimated models were given). We refer the reader to Example 1 for a usage example and exemplary output from the `compare_models` function.

Predicting choices and response times. Choices and RTs can be predicted with the GLAM by the use of the `predict` method:

```
model.predict(n_repeats=50)
```

For each trial of the dataset that is attached to the model instance, this method predicts a choice and RT according to Eq (10), using the previously determined MAP parameter estimates. To obtain a stable estimate of the GLAM's predictions, as well as the noise contained within them, it is recommended to repeat every trial multiple times during the prediction. The number of trial repeats can be specified with the `n_repeats` argument. After the prediction is completed, the predicted data can be accessed with the `prediction` attribute of the model.

Results

Example 1: Individual level data & model comparison

Our first example is based on the study by [21]. Here, the authors study the association between gaze allocation and choice behaviour on the level of the individual. In particular, they explore whether (1) gaze biases are present on the individual level and (2) the strength of this association varies between individuals. In this example, we replicate this type of individual model-based analysis, including parameter estimation, comparison between multiple model variants, and out-of-sample prediction of choice and RT data.

Simulating data. First, we simulate a dataset containing 30 subjects, each performing 300 simple value-based choice trials. We assume that in each trial participants are asked to choose the item that they like most out of a set of three presented alternatives (e.g., snack food items;

similar to the task described in [8]). While participants perform the task, their eye movements, choices and RTs are measured. Before completing the choice trials, participants were asked to indicate their liking rating for each of the items used in the choice task on a liking rating scale between 1 and 10 (with 10 indicating strong liking and 1 indicating little liking). The resulting dataset contains a liking value for each item in a trial, the participants' choice and RT, as well as the participant's gaze towards each item in a trial (describing the fraction of trial time that the participant spent looking at each item in the choice set).

To simulate individuals' response behaviour, we utilize the parameter estimates that were obtained by [21] for the individuals in the three item choice dataset by [8] (see S1 Fig). Importantly, we assume that ten individuals do not exhibit a gaze bias, meaning that their choices are independent of the time that they spend looking at each item. To this end, we set the γ value of ten randomly selected individuals to 1. We further assume that individuals' gaze is distributed randomly with respect to the values of the items in a choice set. An overview of the generating parameter estimates is given in S3 Fig.

We first instantiate a GLAM model instance using `gb.GLAM()` and then use its `simulate_group` method. This method requires us to specify whether the individuals of the group are either simulated individually (and thereby independent of one another) or as part of a group with hierarchical parameter structure (where the individual model parameters are drawn from a group distribution, see below). For the former, the generating model parameters (indicated in the following as `gen_parameters`) are provided as a dictionary, containing a list of the individual participant values for each model parameter:

```
import glambox as gb
import numpy as np
glam = gb.GLAM()
no_bias_subjects = np.random.choice(a=gen_parameters
['gamma'].size,
                                    size=10,
                                    replace=False)
gen_parameters['gamma'][no_bias_subjects] = 1
glam.simulate_group(kind='individual',
                    n_individuals=30,
                    n_trials=300,
                    n_items=3,
                    parameters=gen_parameters)
```

As this example is focused on the individual level, we can further create a summary table, describing individuals' response behaviour on three behavioural metrics, using the `aggregate_subject_level_data` function from the `analysis` module. The resulting table contains individuals' mean RT, their probability of choosing the item with the highest item value from a choice set and a behavioural measure of the strength of the association between individuals' gaze allocation and choice behaviour (indicating the mean increase in choice probability for an item that was fixated on longer than the others, after correcting for the influence of the item value on choice behaviour; for further details, see [21]).

```
from glambox.analysis import aggregate_subject_level_data
subject_data_summary = aggregate_subject_level_data
(data=glam.data,
                                                    n_items=3)
```

Exploring the behavioural data. In a first step of our analysis, we explore differences in individuals' response behaviour. To this end, we plot the distributions of individuals' scores on

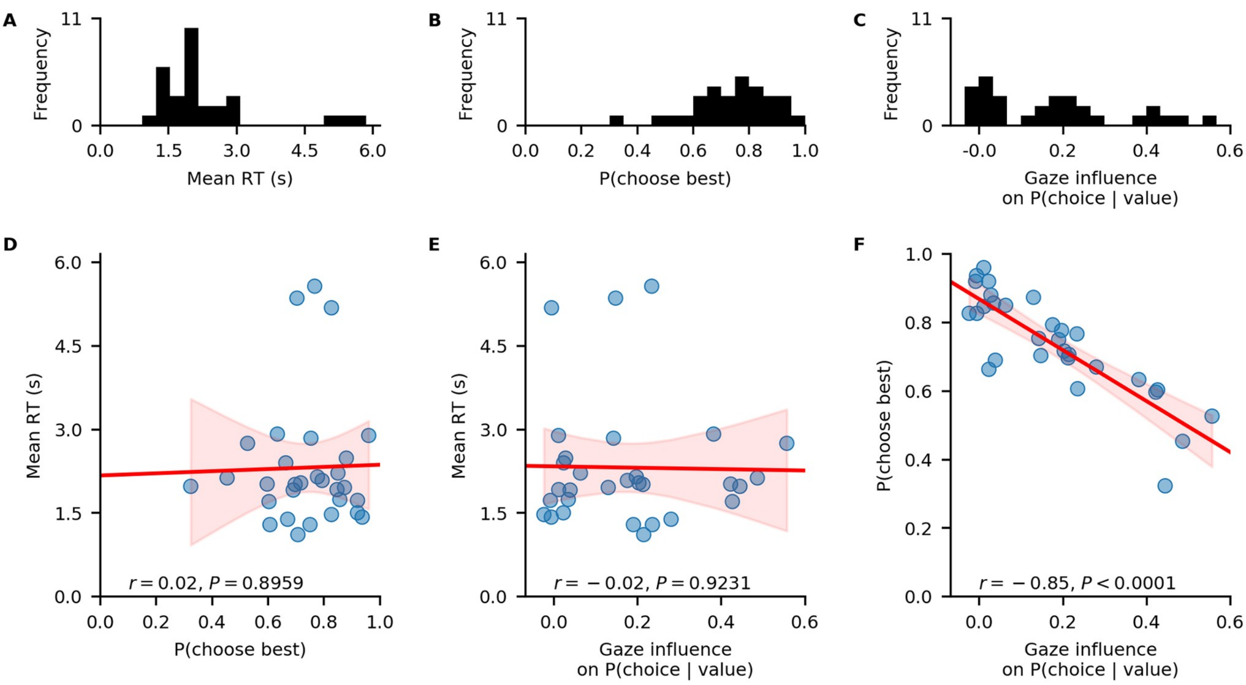


Fig 3. Individual differences in the data. A-C: distributions of individuals' mean RT (A), probability of choosing the highest-valued item in a trial (B), and behavioural influence of gaze allocation on choice behaviour (C). D-F: associations between individuals' probability of choosing the highest-valued item and mean RT (D), individuals' behavioural influence of gaze allocation on choice behaviour and their mean RT (E), individuals' behavioural influence of gaze allocation on choice behaviour and their probability of choosing the highest-valued item (F). Red lines indicate linear regression fits with confidence bands surrounding them. Pearson's *r* coefficients with corresponding *P*-values are reported for each association in D-F.

https://doi.org/10.1371/journal.pone.0226428.g003

the three behavioural metrics, and their associations, using the `plot_behaviour_associations` function implemented in the `plots` module:

```
gb.plots.plot_behaviour_associations(data=data)
```

The resulting plot is displayed in Fig 3 and shows that individuals' probability of choosing the best item, as well as the strength of their behavioural association of gaze and choice, are not associated with their mean RT (Fig 3D and 3E). However, individuals' probability of choosing the best item increases with decreasing strength of the behavioural association of gaze and choice (Fig 3F).

Likelihood-based model comparison. In a second step of our analysis, we want to test whether the response behaviour of each individual is better described by a decision model with or without gaze bias. To this end, we set up the two GLAM variants:

```
glam_bias = gb.GLAM(data=data)
glam_bias.make_model(kind='individual', name='glam_bias')

glam_nobias = gb.GLAM(data=data)
glam_nobias.make_model(kind='individual', gamma_val=1,
name='glam_nobias')
```

For the GLAM variant without gaze bias mechanism, we use the `gamma_val` argument and set it to a value of 1 (fixing γ to 1 for all subjects). We also assign different names to each model with the `name` attribute to better identify them in our subsequent analyses.

Subsequently, we fit both models to the data of each individual and compare their fit by means of the Widely Applicable Information Criterion (WAIC; [34]):

```
glam_bias.fit(method='MCMC',
              tune=5000,
              draws=5000,
              chains=4)

glam_nobias.fit(method='MCMC',
                tune=5000,
                draws=5000,
                chains=4)
```

The fit method defaults to Metropolis-Hastings MCMC sampling (for methodological details, see Methods Section). The `draws` argument sets the number of samples to be drawn. This excludes the tuning (or burn-in) samples, which can be set with the `tune` argument. In addition, the `fit` method accepts the same keyword arguments as the PyMC3 sample function, which it wraps (see the PyMC3 documentation for additional details). The `chains` argument sets the number of MCMC traces (it defaults to four and should be set to at least two, in order to allow convergence diagnostics).

After convergence has been established for all parameter traces (for details on the suggested convergence criteria, see Methods), we perform a model comparison on the individual level, using the `compare_models` function from the `analysis` (see Basic Usage: Comparing model variants):

```
comparison_df = gb.analysis.compare_models(models=[glam_bias,
glam_nobias],
                                           ic='WAIC')
```

The resulting table (shown in Table 2) can be used to identify the best fitting model (indicated by the lowest WAIC score) per individual.

With this comparison, we are able to identify those participants whose response behaviour matches the assumption of gaze-biased evidence accumulation. In particular, we find that we accurately recover whether an individual has a gaze bias or not for 29 out of 30 individuals.

Looking at the individual parameter estimates (defined as MAP of the posterior distributions), we find that the individually fitted γ values (Fig 4A) cover a wide range between -0.8 and 1, indicating strong variability in the strength of individuals' gaze bias. We also find that γ estimates have a strong negative correlation with individuals' scores on the behavioural gaze bias measure (Fig 4B).

Out-of-sample prediction. We have identified those participants whose response behaviour is better described by a GLAM variant with gaze-bias than one without. Yet, this analysis does not indicate whether the GLAM is a good model of individuals' response behaviour on an absolute level. To test this, we perform an out-of-sample prediction exercise.

Table 2. Output from `compare_models` function for the first two subjects.

subject	model	WAIC	pWAIC	dWAIC	weight	SE	dSE	var_warn
0	glam_bias	523.6	5.75	0	0.94	50.25	0	0
0	glam_nobias	645.09	3.64	121.49	0.06	44.15	23.56	0
1	glam_bias	1097.86	3.69	0	1	40.32	0	0
1	glam_nobias	1185.02	2.85	87.16	0	38.22	18	0

https://doi.org/10.1371/journal.pone.0226428.t002

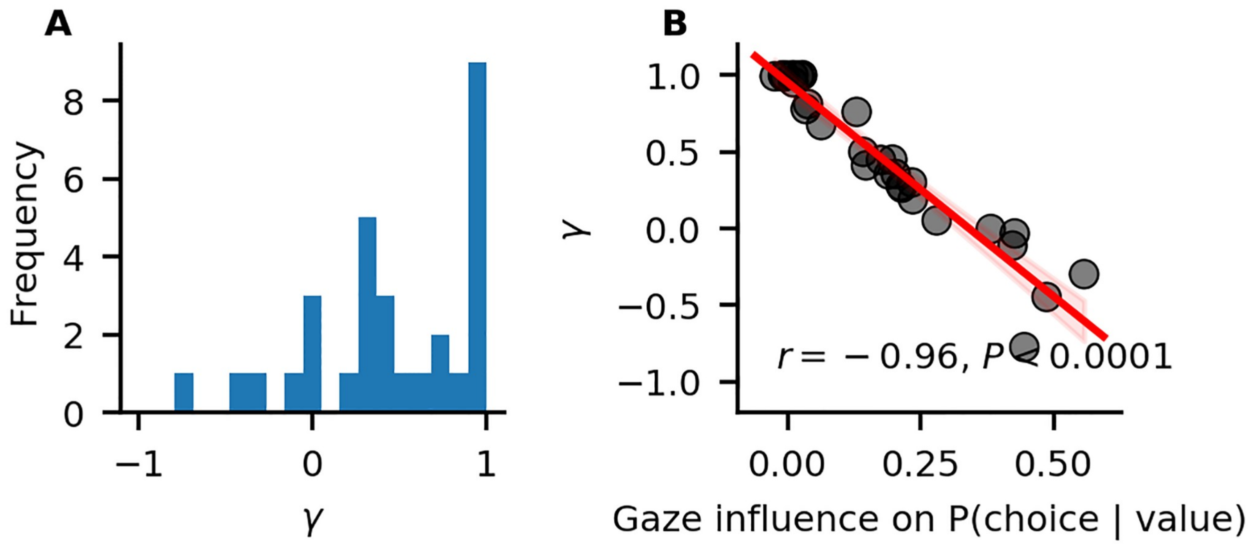


Fig 4. Individual differences in the strength of the association of gaze allocation and choice behaviour. A: Distribution of γ estimates resulting from the in-sample individual model fits. B: Association of γ estimates and individuals' values on the behavioural gaze bias measure. The red line indicates a linear regression fit, with surrounding 95% confidence bands. Pearson's r correlation with P-value is given.

https://doi.org/10.1371/journal.pone.0226428.g004

We divide the data of each subject into even- and odd-numbered experiment trials and use the data of the even-numbered trials to fit both GLAM variants:

```
glam_bias.exchange_data(data_even)
glam_bias.fit(method='MCMC',
              tune=5000,
              draws=5000,
              chains=4)

glam_nobias.exchange_data(data_even)
glam_nobias.fit(method='MCMC',
              tune=5000,
              draws=5000,
              chains=4)
```

Subsequently, we evaluate the performance of both models in predicting individuals' response behaviour using the MAP estimates and item value and gaze data from the odd-numbered trials. To predict response behaviour for the odd-numbered trials, we use the `predict` method. We repeat every trial 50 times in the prediction (as specified through the `n_repeats` argument) to obtain a stable pattern of predictions:

```
glam_bias.exchange_data(data_odd)
glam_bias.predict(n_repeats=50)

glam_nobias.exchange_data(data_odd)
glam_nobias.predict(n_repeats=50)
```

Lastly, to determine the absolute fit of both model variants to the data, we plot the individually predicted against the individually observed data on all three behavioural metrics. To do this, we use the `plot_individual_fit` function of the `plots` module. This function takes as input the observed data, as well as a list of the predictions of all model variants that ought to be compared. The argument `prediction_labels` specifies the naming used for each model in the resulting figure. For each model variant, the function creates a row of panels, plotting the observed against the predicted data:

```
from glambox.plots import plot_individual_fit
plot_individual_fit(observed = data_odd,
                    predictions=[glam_bias.prediction,
                                 glam_nobias.prediction],
                    prediction_labels=['gaze-bias',
                                       'no gaze-bias'])
```

The resulting plot is displayed in Fig 5. We find that both model variants perform well in capturing individuals' RTs and probability of choosing the best item (Fig 5A, 5D, 5B and 5E). Importantly, only the GLAM variant with gaze bias is able to also recover the strength of the association between individuals' choice behaviour and gaze allocation (Fig 5C).

Conclusion. GLAMbox provides an easy-to-use tool to test the presence (and variability) of gaze biases on the individual level. With GLAMbox, we can easily fit the GLAM to individual participant data, compare different model variants and predict individuals' response behaviour. It also provides a set of analysis functions to explore behavioural differences between individuals and to compare the fit of different model variants to observed response behaviour.

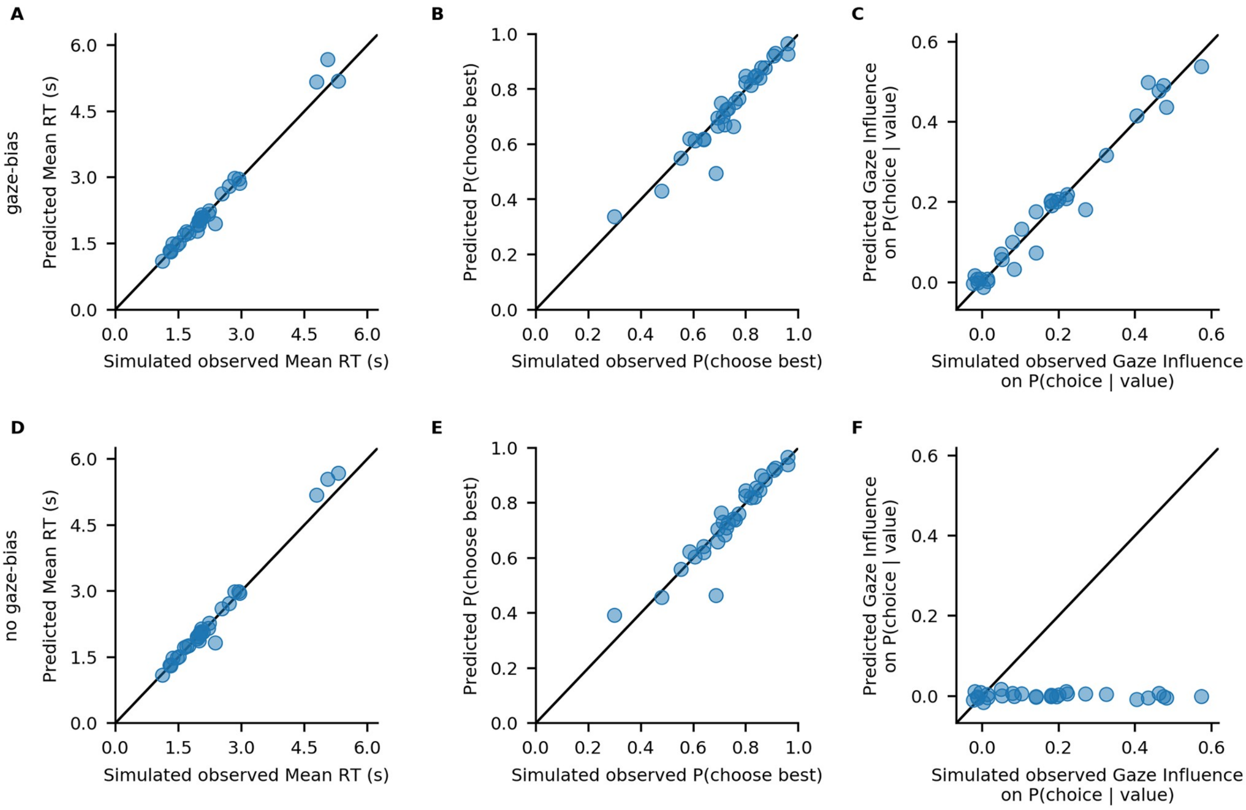


Fig 5. Out-of-sample model fits. Comparison of individuals' simulated observed response behaviour with the out-of-sample predictions of a GLAM variant with (A-C) and without gaze bias (D-F): Individuals' mean RT (A, D), probability of choosing the best item (B, E), and influence of gaze allocation on choice probability (C, F). Points indicate individual participant means.

https://doi.org/10.1371/journal.pone.0226428.g005

Example 2: Hierarchical parameter estimation in cases with limited data

In some research settings, the total amount of data one can collect per individual is limited, conflicting with the large amounts of data required to obtain reliable and precise individual parameter estimates from diffusion models [35, 36]. Hierarchical modeling can offer a solution to this problem. Here, each individual's parameter estimates are assumed to be drawn from a group level distribution. Thereby, during parameter estimation, individual parameter estimates are informed by the data of the entire group. This can greatly improve parameter estimation, especially in the face of limited amounts of data [25, 37]. In this example, we will simulate a clinical application setting, in which different patient groups are to be compared on the strengths of their gaze biases, during a simple value-based choice task that includes eye tracking. It is reasonable to assume that the amount of data that can be collected in such a setting is limited on at least two accounts:

1. The number of patients available for the experiment might be low
2. The number of trials that can be performed by each participant might be low, for clinical reasons (e.g., patients feel exhausted more quickly, time to perform tests is limited, etc.)

Therefore, we simulate a dataset with a low number of individuals within each group (between 5 and 15 per group), and a low number of trials per participant (50 trials). We then estimate model parameters in a hierarchical fashion, and compare the group level gaze bias parameter estimates between groups.

Simulating data. We simulate data of three patient groups ($N_1 = 5$, $N_2 = 10$, $N_3 = 15$), with 50 trials per individual, in a simple three item value-based choice task, where participants are instructed to simply choose the item they like the best. These numbers are roughly based on a recent clinical study on the role of the prefrontal cortex in fixation-dependent value representations [13]. Here, the authors found no systematic differences between frontal lobe patients and controls on integration speed or the decision threshold, controlling speed-accuracy trade-offs. Therefore, in our example we only let the gaze bias parameter γ differ systematically between the groups, with means of $\gamma_1 = 0.7$ (weak gaze bias), $\gamma_2 = 0.1$ (moderate gaze bias) and $\gamma_3 = -0.5$ (strong gaze bias), respectively. We do not assume any other systematic differences between the groups and sample all other model parameters from the estimates obtained from fitting the model to the data of [8] (for an overview of the generating parameters, see S4 Fig).

Behavioural differences between the three groups are plotted in Fig 6, using the `plot_behaviour_aggregate` function from the `plots` module. Group-level

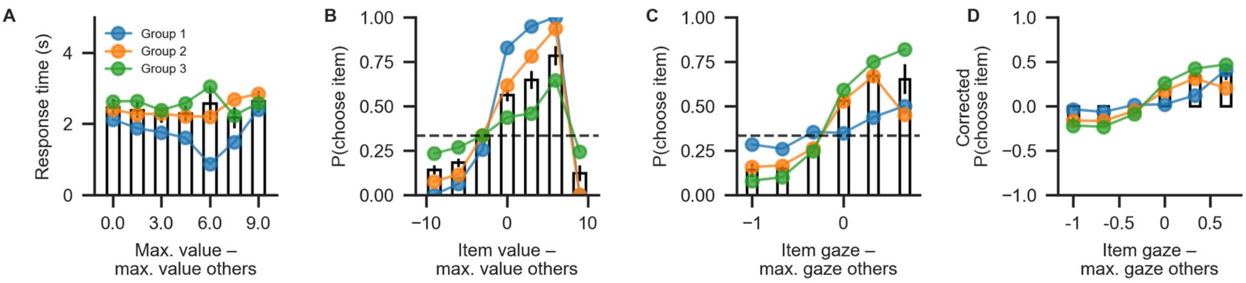


Fig 6. Aggregate view of the simulated data for Example 2. (A) Mean RT binned by trial difficulty (the difference between the highest item value in a choice set and the maximum value of all other items). (B) The probability that an item is chosen based on its relative value (the difference of the item's value and the maximum value of all other items in the choice set). (C) The probability of choosing an item based on its relative gaze (the difference between the gaze towards this item and the maximum gaze towards a different item). (D) The probability of choosing an item based on its relative gaze, when correcting for the influence of its value. Bars correspond to the pooled data, while coloured lines indicate individual groups.

https://doi.org/10.1371/journal.pone.0226428.g006

summary tables can be created using the `aggregate_group_level_data` from the `analysis` module. Even though the groups only differ in the gaze bias parameter, they also exhibit differences in RT (Group 1 mean ± s.d. = 1.96 ± 0.33 s, Group 2 mean ± s.d. = 2.38 ± 1.4 s; Group 3 mean ± s.d. = 2.59 ± 1.26 ms; Fig 6A) and choice accuracy (Group 1 mean ± s.d. = 0.88 ± 0.06, Group 2 mean ± s.d. = 0.71 ± 0.07, Group 3 mean ± s.d. = 0.50 ± 0.16; Fig 6B). As is to be expected, we can also observe behavioural differences in gaze influence measure (Group 1 mean ± s.d. = 0.08 ± 0.07, Group 2 mean ± s.d. = 0.26 ± 0.11, Group 3 mean ± s.d. = 0.38 ± 0.11; Fig 6C and 6D, where the choices of Group 3 are driven by gaze more than those of the other groups.

Building the hierarchical model. When specifying the hierarchical model, we allow all model parameters to differ between the three groups. This way, we will subsequently be able to address the question whether individuals from different groups differ on one or more model parameters (including the gaze bias parameter γ, which we are mainly interested in here). As for the individual models, we first initialize the model object using the GLAM class and supply it with the behavioural data using the `data` argument. Here, we set the model kind to `'hierarchical'` (in contrast to `'individual'`). Further, we specify that each model parameter can vary between groups (referring to a `'group'` variable in the data):

```
hglam = gb.GLAM(data=data)
hglam.make_model(kind='hierarchical',
                 depends_on = dict(v='group',
                                   gamma='group',
                                   s='group',
                                   tau='group'))
```

In this model, each parameter is set up hierarchically within each group, so that individual estimates are informed by other individuals in that group. If the researcher does not expect group differences on a parameter, this parameter can simply be omitted from the `depends_on` dictionary. The resulting model would then have a hierarchical setup of this parameter across groups, so that individual parameter estimates were informed by all other individuals (not only those in the same group).

Parameter estimation with MCMC. After the model is built, the next step is to perform statistical inference over its parameters. As we have done with the individual models, we can use MCMC to approximate the parameters' posterior distributions (see Methods for details). Due to the more complex structure and drastically increased number of parameters, the chains from the hierarchical model usually have higher levels autocorrelation. To still obtain a reasonable number of effective samples [32], we increase the number of tuning- and draw steps:

```
hglam.fit(method='MCMC',
          draws=20000,
          tune=20000,
          chains=4)
```

Evaluating parameter estimates, interpreting results. After sampling is finished, and the chains were checked for convergence, we can turn back to the research question: Do the groups differ with respect to their gaze biases? Questions about differences between group-level parameters can be addressed by inspecting their posterior distributions. For example, the probability that the mean $\gamma_{1,\mu}$ for Group 1 is larger than the mean $\gamma_{2,\mu}$ of Group 2 is given by the proportion of posterior samples in which this was the case.

GLAMbox includes a `compare_parameters` function that plots posterior distributions of group level parameters. Additionally, the user can specify a list of comparisons between groups or conditions. If comparisons are specified, the posterior distributions of their difference and corresponding relevant statistics are added to the figure:

```
from glambox.plots import compare_parameters
parameters = ['v', 'gamma', 's', 'tau']
comparisons = [('group1', 'group2'),
               ('group1', 'group3'),
               ('group2', 'group3')]
compare_parameters(model=hglam,
                   parameters=parameters,
                   comparisons=comparisons)
```

With the resulting plot (Fig 7), the researcher can infer that the groups did not differ with respect to their mean velocity parameters $v_{i,\mu}$ (top row, pairwise comparisons), mean accumulation noise $\sigma_{i,\mu}$ (third row), or scaling parameters $\tau_{i,\mu}$. The groups differ, however, in the strength of their mean gaze bias $\gamma_{i,\mu}$ (second row): All differences between the groups were

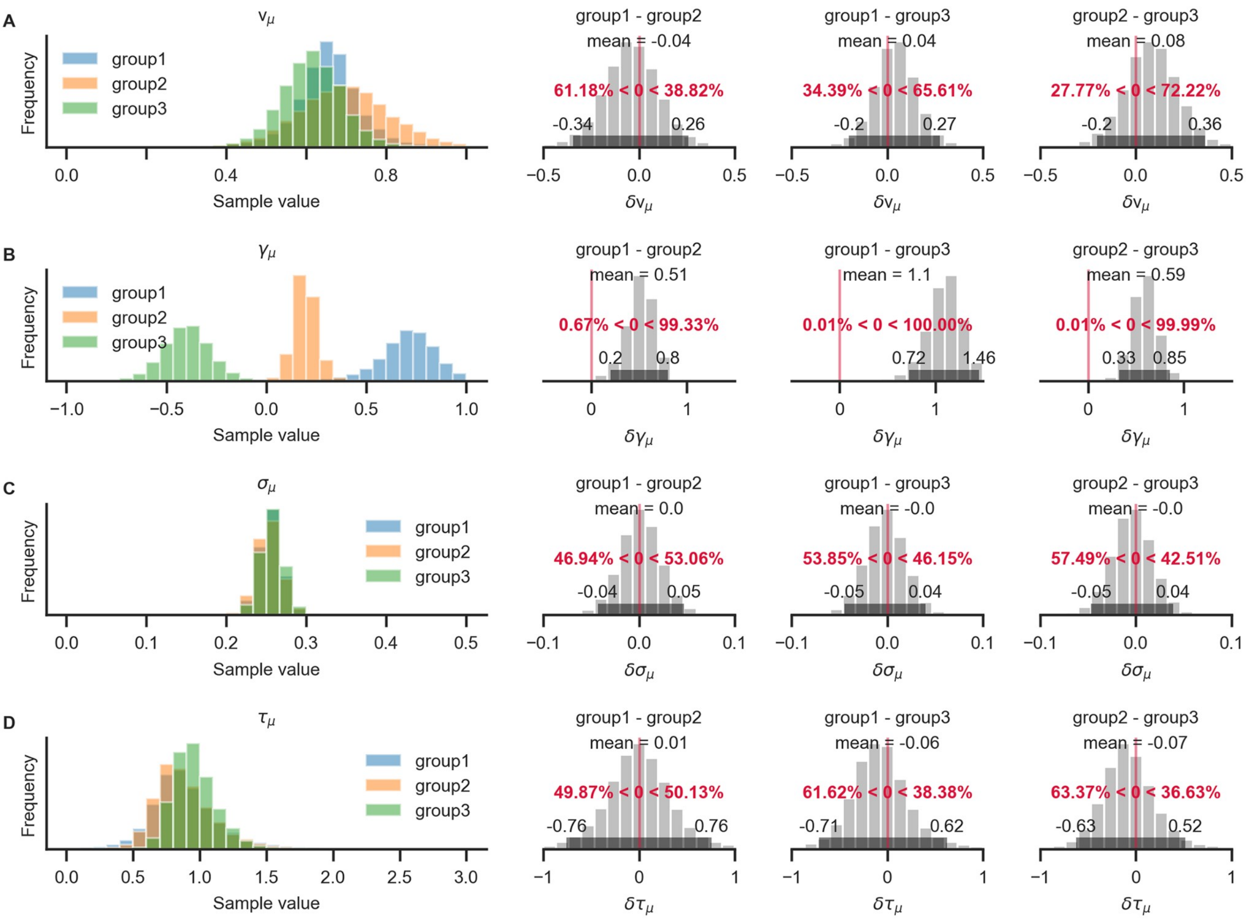


Fig 7. Pairwise comparison of posterior group-level parameter estimates between groups. Each row corresponds to one model parameter. The leftmost column shows the estimated posterior distributions for each parameter and group. Pairwise differences between the group posterior distributions are shown in all other columns. For each posterior distribution of the difference, the mean and 95% HPD are indicated, as well as the proportion of samples below and above zero (in red). All three groups differ on the γ parameter (row B). No evidence for differences on any of the other model parameters is found (the 95% HPD of the pairwise differences between groups all include zero).

https://doi.org/10.1371/journal.pone.0226428.g007

statistically meaningful (as inferred by the fact that the corresponding 95% HPD did not contain zero; second row, columns 2-4).

Conclusion. When faced with limited data, GLAMbox allows users to easily build and estimate hierarchical GLAM variants, including conditional dependencies of model parameters. The Bayesian inference framework allows the researcher to answer relevant questions in a straightforward fashion. To this end, GLAMbox provides basic functions for computation and visualization.

Example 3: Parameter recovery

When performing model-based analyses of behaviour that include the interpretation of parameter estimates, or comparisons of parameter estimates between groups or conditions, the researcher should be confident that the model's parameters are actually identifiable. In particular, the researcher needs to be confident that the set of estimated parameters unambiguously describes the observed data better than any other set of parameters. A straightforward way of testing this is to perform a parameter recovery: The general intuition of a parameter recovery analysis is to first generate a synthetic dataset from a model using a set of known parameters, and then fitting the model to the synthetic data. Finally, the estimated parameters can be compared to the known generating parameters. If they match to a satisfying degree, the parameters were recovered successfully. Previous analyses have already indicated that the GLAM's parameters can be recovered to a satisfying degree [21]. Yet, the ability to identify a given set of parameters always depends on the specific features of a given dataset. The most obvious feature of a dataset that influences recoverability of model parameters is the number of data points included. Usually this quantity refers to the number of trials that participants performed. For hierarchical models, the precision of group-level estimates also depends on the number of individuals per group. Additional features that vary between datasets and that could influence parameter estimation are the observed distribution of gaze, the distribution of item values or the number of items in each trial. For this reason, it is recommended to test whether the estimated parameters of a model can be recovered in the context of a specific dataset. slac To demonstrate the procedure of a basic parameter recovery analysis using GLAMbox, suppose we have collected and loaded a dataset called `data`. In the first step, we perform parameter estimation as in the previous examples:

```
glam = gb.GLAM(data=data)
glam.make_model(kind='individual')
glam.fit(method='MCMC',
         draws=5000,
         tune=5000,
         chains=4)
```

The next step is to create a synthetic, model-generated dataset using the model parameters estimated from the empirical data, together with the empirically observed stimulus and gaze data using the `predict` method. Setting `n_repeats` to 1 results in a dataset of the same size as the observed one:

```
glam.predict(n_repeats=1)
synthetic = glam.prediction
```

The synthetic dataset should resemble the empirically observed data closely. If there are major discrepancies between the synthetic and observed data, this indicates that GLAM might not be a good candidate model for the data at hand. Next, we create a new model instance, attach the synthetic data, build a model and re-estimate its parameters:

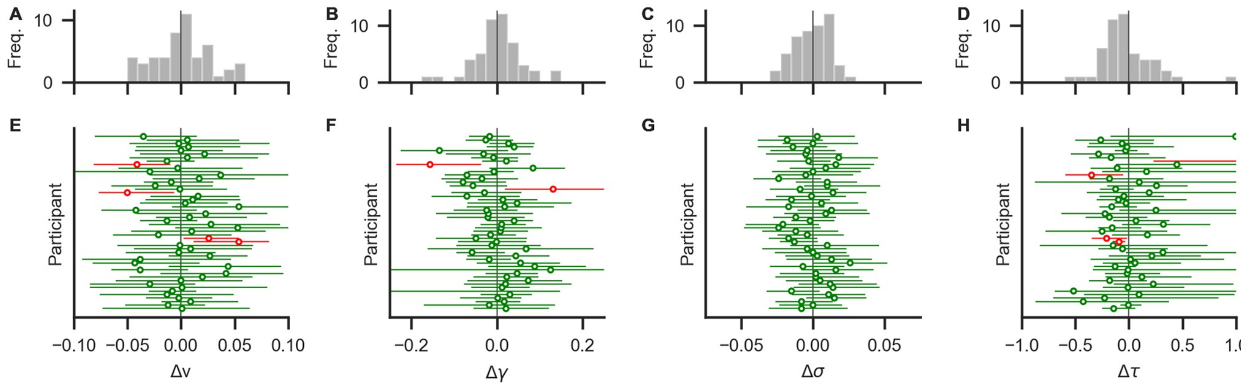


Fig 8. Results from a basic parameter recovery. The lower row (E-H) shows deviations between known generating parameter values and recovered MAP estimates (circles) and their 95% HPDs (horizontal error bars) for each participant. Green (red) colour indicates that the true value is within (outside) the 95% HPD. Most parameters were recovered with small deviations. Panels A-D show distributions of deviations across individuals. Distributions are mostly centered around zero, indicating no systematic under- or overestimation (bias) across individuals.

https://doi.org/10.1371/journal.pone.0226428.g008

```
glam_rec = gb.GLAM(data=synthetic)
glam_rec.make_model(kind='individual')
glam_rec.fit(method='MCMC',
             draws=5000,
             tune=5000,
             chains=4)
```

Finally, the recovered and generating parameters can be compared. If the recovered parameters do not match the generating parameters, the parameters cannot be identified given this specific dataset. In this case, parameter estimates should not be interpreted.

If, on the other hand, generating and recovered parameters do align, the parameters have been recovered successfully. This indicates that the model's parameters can be identified unambiguously given the general characteristics of the dataset and thereby increases confidence that the parameters obtained from the empirical data are valid and can be interpreted.

Here, all parameters could be recovered as illustrated in Fig 8. For most individuals, the MAP estimates and their 95% HPDs are close to the known generating parameters. Across individuals, no systematic biases in the estimation can be identified.

Conclusion. In this example, we demonstrated how to perform a basic parameter recovery for a given dataset. When successful, this increases confidence that the parameters can be identified with the given dataset.

Discussion

Researchers have recently started to systematically investigate the role of visual gaze in the decision making process. By now, it is established that eye movements do not merely serve to sample information that is then processed independently to produce a choice, but that they are actively involved in the construction of preferences [2, 4, 6–8, 10, 14, 15, 21, 38]. The dominant theoretical perspective is that evidence accumulation in favor of each option is modulated by gaze allocation, so that accumulation for non-fixated options is attenuated. This mechanism is formally specified in various models of gaze-dependent decision making, such as the attentional Drift Diffusion Model (aDDM; [7, 8]) and the conceptually related Gaze-weighted Linear Accumulator Model (GLAM; [21]). In contrast to analyses based on behavioural and eye tracking data alone, these models can act as analytical tools that enable researchers to address questions regarding specific mechanisms in the decision process, like the gaze bias.

The goal of GLAM is to provide a model-based estimate of the gaze bias on the level of an individual (as indicated by GLAM's γ parameter), in choice situations involving more than two choice alternatives. To estimate the gaze bias, GLAM describes the decision process in the form of a linear stochastic race and aggregates over the specific sequence of fixations during the decision process (by only utilizing the fraction of the decision time that each item was looked at). These two characteristics distinguish the GLAM from other existing approaches of obtaining an estimate of individuals' gaze bias:

First, the GLAM is focused on quantifying the gaze bias on the individual level. It does not capture dynamics of the decision process on the level of single fixations. If these fine-grained dynamics are of interest to the researcher, the aDDM can be used. Here, the fixation-dependent changes in evidence accumulation rates throughout the trial are not averaged out. Keeping this level of detail, however, comes at a cost: Fitting the aDDM relies on extensive model simulations (including a simulation of the fixation process; for a more detailed discussion see [21]). The GLAM, on the other hand, aggregates over the fixation-dependent changes in the accumulator's drift rate in order to simplify the estimation process of the gaze bias.

Second, the GLAM directly applies to choice situations involving more than two choice alternatives. While the GLAM has been shown to also capture individuals' gaze bias and choice behaviour well in two-alternative choice situations [21], there exist other computational approaches that can estimate the gaze bias of an individual in binary decisions: If response times are of interest to the researcher, the gaze bias can be estimated in the form of a gaze-weighted DDM (see for example [2, 18]). Similar to the GLAM, this approach also aggregates over the dynamics of the fixation process within a trial, by only utilizing the fraction of trial time that each item was looked at. In contrast to the GLAM, however, gaze-weighted DDM approaches describe the decision process in the form of a single accumulator that evolves between two decision bounds (each representing one of the two choice alternatives). For two-alternative choice scenarios, where response times are not of interest to the researcher, Smith and colleagues [39] proposed a method of estimating the aDDM gaze-bias parameter through a random utility model. Here, the gaze bias can be estimated in a simple logit model.

Even though the advantages of applying these types of models are apparent, their use is often limited by their complexity and the high cost of implementing, validating and optimizing them. Further, there are only few off-the-shelf solutions researchers can turn to, if they want to perform model-based analyses of gaze-dependent choice data, particularly for choice settings involving more than two alternatives.

With GLAMbox, we present a Python-based toolbox, built on top of PyMC3, that allows researchers to perform model-based analyses of gaze-bias effects in decision making easily. We have provided step-by-step instructions and code to perform essential modeling analyses using the GLAM. These entail application of the GLAM to individual and group-level data, specification of parameter dependencies for both within- and between-subject designs, (hierarchical) Bayesian parameter estimation, comparisons between multiple model variants, out-of-sample prediction of choice and RT data, data visualization, Bayesian comparison of posterior parameter estimates between conditions, and parameter recovery. We hope that GLAMbox will make studying the association between gaze allocation and choice behaviour more accessible. We also hope that the resulting findings will ultimately help us better understand this association, its inter-individual variability and link to brain activity.

Supporting information

S1 Fig. Distribution of individual parameter estimates from four datasets analysed in Thomas et al. (2019). The top row contains distributions of parameter estimates across datasets. Subsequent rows show distributions per dataset: Krajbich et al. (2010; blue), Krajbich & Rangel (2011; orange), Experiment 2 from Folke et al. (2017; green) and Experiment 1 from Tavares et al. (2017; red).
(TIFF)

S2 Fig. Illustration of hyperpriors. Different hyperpriors based on group-averaged parameter values were obtained from fitting the model to four different datasets (Folke et al., 2017; Krajbich et al., 2010; Krajbich & Rangel, 2011; Tavares et al., 2017; see S1 Table and S1 Fig). Panels show prior distributions on group level mean (upper row) and standard deviation (lower row) for each model parameter (columns; from left to right: ν, γ, σ, τ). Observed group level estimates from the four datasets are indicated as red ticks in each panel. Blue, orange and green lines represent prior distributions with increasing levels of vagueness f. They are constructed as normal distributions with mean equal to the mean of the observed group level parameters across datasets (M), and standard deviation equal to f times the observed standard deviation across datasets (SD). Higher values of f correspond to wider, less informative priors. Prior distributions are further bounded between sensible limits. The user can specify the factor f during specification of hierarchical models. By default, hyperpriors with $f = 10$ (orange lines) are used.
(TIFF)

S3 Fig. Distribution of individual generating GLAM parameters of Example 1. Colours indicate whether a subject was simulated with or without gaze bias.
(TIFF)

S4 Fig. Distributions of data-generating parameters for the three groups in Example 2. The top row shows distributions pooled across groups. The bottom three rows show distributions per group. Note that the groups do not differ systematically with respect to the velocity parameter ν, the noise parameter σ, or the scaling parameter τ (first, second and last column; even though there is some variability between individuals). The groups differ, however, on the gaze bias parameter γ (third column): Group 1 only has a weak gaze bias (large γ), group 2 has a medium strong gaze bias (smaller γ), and group 3 has a very strong gaze bias (even smaller, negative γ).
(TIFF)

S1 Table. Description of individual parameter estimates from four datasets analysed in Thomas et al. (2019). The datasets are originally from Folke et al., 2017 (Experiment 2); Krajbich et al., 2010; Krajbich & Rangel, 2011 and Tavares et al., 2017 (Experiment 1).
(PDF)

Author Contributions

Conceptualization: Felix Molter, Armin W. Thomas, Hauke R. Heekeren, Peter N. C. Mohr.

Formal analysis: Felix Molter, Armin W. Thomas.

Software: Felix Molter, Armin W. Thomas.

Supervision: Hauke R. Heekeren, Peter N. C. Mohr.

Visualization: Felix Molter, Armin W. Thomas.

Writing – original draft: Felix Molter, Armin W. Thomas.

Writing – review & editing: Felix Molter, Armin W. Thomas, Hauke R. Heekeren, Peter N. C. Mohr.

References

1. Armel KC, Beaumel A, Rangel A. Biasing simple choices by manipulating relative visual attention. Judgment and Decision Making. 2008; 3:396–403.
2. Cavanagh JF, Wiecki TV, Kochar A, Frank MJ. Eye tracking and pupillometry are indicators of dissociable latent decision processes. Journal of Experimental Psychology: General. 2014; 143(4):1476–1488. https://doi.org/10.1037/a0035813
3. Fiedler S, Glöckner A. The Dynamics of Decision Making in Risky Choice: An Eye-Tracking Analysis. Frontiers in Psychology. 2012; 3. https://doi.org/10.3389/fpsyg.2012.00335
4. Folke T, Jacobsen C, Fleming SM, De Martino B. Explicit representation of confidence informs future value-based decisions. Nature Human Behaviour. 2017; 1(1):0002. https://doi.org/10.1038/s41562-016-0002
5. Glöckner A, Herbold AK. An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. Journal of Behavioral Decision Making. 2011; 24(1):71–98. https://doi.org/10.1002/bdm.684
6. Konovalov A, Krajbich I. Gaze data reveal distinct choice processes underlying model-based and model-free reinforcement learning. Nature Communications. 2016; 7:12438. https://doi.org/10.1038/ncomms12438 PMID: 27511383
7. Krajbich I, Armel C, Rangel A. Visual fixations and the computation and comparison of value in simple choice. Nature Neuroscience. 2010; 13(10):1292–1298. https://doi.org/10.1038/nn.2635 PMID: 20835253
8. Krajbich I, Rangel A. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. Proceedings of the National Academy of Sciences. 2011; 108(33):13852–13857. https://doi.org/10.1073/pnas.1101328108
9. Pärnamets P, Johansson P, Hall L, Balkenius C, Spivey MJ, Richardson DC. Biasing moral decisions by exploiting the dynamics of eye gaze. Proceedings of the National Academy of Sciences. 2015; 112(13):4170–4175. https://doi.org/10.1073/pnas.1415250112
10. Shimojo S, Simion C, Shimojo E, Scheier C. Gaze bias both reflects and influences preference. Nature Neuroscience. 2003; 6(12):1317–1322. https://doi.org/10.1038/nn1150 PMID: 14608360
11. Stewart N, Gächter S, Noguchi T, Mullett TL. Eye Movements in Strategic Choice. Journal of Behavioral Decision Making. 2016; 29(2-3):137–156. https://doi.org/10.1002/bdm.1901 PMID: 27513881
12. Stewart N, Hermens F, Matthews WJ. Eye Movements in Risky Choice. Journal of Behavioral Decision Making. 2016; 29(2-3):116–136. https://doi.org/10.1002/bdm.1854 PMID: 27522985
13. Vaidya AR, Fellows LK. Testing necessary regional frontal contributions to value assessment and fixation-based updating. Nature Communications. 2015; 6: 10120. https://doi.org/10.1038/ncomms10120
14. Tavares G, Perona P, Rangel A. The Attentional Drift Diffusion Model of Simple Perceptual Decision-Making. Frontiers in Neuroscience. 2017; 11. https://doi.org/10.3389/fnins.2017.00468 PMID: 28894413
15. Ashby NJS, Jekel M, Dickert S, Glöckner A. Finding the right fit: A comparison of process assumptions underlying popular drift-diffusion models. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2016; 42(12):1982–1993. https://doi.org/10.1037/xlm0000279 PMID: 27336785
16. Fisher G. An attentional drift diffusion model over binary-attribute choice. Cognition. 2017; 168:34–45. https://doi.org/10.1016/j.cognition.2017.06.007 PMID: 28646751
17. Krajbich I, Lu D, Camerer C, Rangel A. The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions. Frontiers in Psychology. 2012; 3. https://doi.org/10.3389/fpsyg.2012.00193
18. Lopez-Persem A, Domenech P, Pessiglione M. How prior preferences determine decision-making frames and biases in the human brain. eLife. 2016; 5:e20317. https://doi.org/10.7554/eLife.20317 PMID: 27864918
19. Trueblood JS, Brown SD, Heathcote A, Busemeyer JR. Not Just for Consumers: Context Effects Are Fundamental to Decision Making. Psychological Science. 2013; 24(6):901–908. https://doi.org/10.1177/0956797612464241 PMID: 23610134
20. Smith SM, Krajbich I. Attention and choice across domains. Journal of Experimental Psychology: General. 2018; 147(12):1810–1826. https://doi.org/10.1037/xge0000482

21. Thomas AW, Molter F, Krajbich I, Heekeren HR, Mohr PNC. Gaze bias differences capture individual choice behaviour. Nature Human Behaviour. 2019; 3(6):625–635. https://doi.org/10.1038/s41562-019-0584-8 PMID: 30988476

22. Cavanagh SE, Malalasekera WMN, Miranda B, Hunt LT, Kennerley SW. Visual Fixation Patterns during Economic Choice Reflect Covert Valuation Processes That Emerge with Learning. Proceedings of the National Academy of Sciences. 2019; p. 201906662.

23. Gwinn R, Krajbich I. Attitudes and Attention. Journal of Experimental Social Psychology. 2020; 86:103892. https://doi.org/10.1016/j.jesp.2019.103892

24. Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD. The Physics of Optimal Decision Making: A Formal Analysis of Models of Performance in Two-Alternative Forced-Choice Tasks. Psychological review. 2006; 113(4):700. https://doi.org/10.1037/0033-295X.113.4.700 PMID: 17014301

25. Wiecki TV, Sofer I, Frank MJ. HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. Frontiers in Neuroinformatics. 2013; 7. https://doi.org/10.3389/fninf.2013.00014 PMID: 23935581

26. Ratcliff R, Tuerlinckx F. Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. Psychonomic Bulletin & Review. 2002; 9(3):438–481. https://doi.org/10.3758/BF03196302

27. Salvatier J, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. PeerJ Computer Science. 2016; 2:e55. https://doi.org/10.7717/peerj-cs.55

28. McKinney W. Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference. vol. 445. Austin, TX; 2010. p. 51–56.

29. Voss A, Voss J. Fast-dm: A free program for efficient diffusion model analysis. Behavior Research Methods. 2007; 39(4):767–775. https://doi.org/10.3758/bf03192967 PMID: 18183889

30. Gamerman D, Lopes HF. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition. Chapman and Hall/CRC; 2006.

31. Gelman A, Shirley K. Inference from Simulations and Monitoring Convergence. In: Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC; 2011. p. 189–200.

32. Kruschke J. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Academic Press; 2014.

33. Kucukelbir A, Tran D, Ranganath R, Gelman A, Blei DM. Automatic differentiation variational inference. The Journal of Machine Learning Research. 2017; 18(1):430–474.

34. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing. 2017; 27(5):1413–1432. https://doi.org/10.1007/s11222-016-9696-4

35. Lerche V, Voss A, Nagler M. How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. Behavior Research Methods. 2017; 49(2):513–537. https://doi.org/10.3758/s13428-016-0740-2 PMID: 27287445

36. Voss A, Nagler M, Lerche V. Diffusion Models in Experimental Psychology. Experimental Psychology. 2013; 60(6):385–402. https://doi.org/10.1027/1618-3169/a000218 PMID: 23895923

37. Ratcliff R, Childers R. Individual Differences and Fitting Methods for the Two-Choice Diffusion Model of Decision Making. Decision (Washington, DC). 2015; 2015.

38. Orquin JL, Mueller Loose S. Attention and choice: A review on eye movements in decision making. Acta Psychologica. 2013; 144(1):190–206. https://doi.org/10.1016/j.actpsy.2013.06.003 PMID: 23845447

39. Smith SM, Krajbich I, Webb R. Estimating the dynamic role of attention via random utility. Journal of the Economic Science Association. 2019; p. 1–15.

Molter, F., Thomas, A., Huettel, S. A., Heekeren, H. R., & Mohr, P. N. C. (submitted). Gaze-dependent evidence accumulation predicts multi-alternative risky choice behaviour. *PsyArXiv*. https://doi.org/10.31234/osf.io/x6nbf

Gaze-dependent evidence accumulation predicts multi-alternative risky choice behaviour

Felix Molter[1,2,3,*], Armin W. Thomas[2,4], Scott A. Huettel[5,6], Hauke R. Heekeren[2,7], and Peter N. C. Mohr[1,2,3]

[1]School of Business & Economics, Freie Universität Berlin, Berlin, Germany
[2]Center for Cognitive Neuroscience, Freie Universität Berlin, Berlin, Germany
[3]WZB Berlin Social Science Center, Berlin, Germany
[4]Department of Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany
[5]Center for Cognitive Neuroscience, Duke University, Durham, NC, USA
[6]Department for Psychology and Neuroscience, Duke University, Durham, NC, USA
[7]Department for Education and Psychology, Freie Universität Berlin, Berlin, Germany
[*]E-mail: felixmolter@gmail.com

February 12, 2021

Abstract

Choices are influenced by gaze allocation during deliberation, so that fixating an alternative longer leads to increased probability of choosing it. Gaze-dependent evidence accumulation provides a parsimonious account of choices, response times and gaze-behaviour in many simple decision scenarios. Here, we test whether this framework can also predict more complex context-dependent patterns of choice in a three-alternative risky choice task, where choices and eye movements were subject to attraction and compromise effects. Choices were best described by a gaze-dependent evidence accumulation model, where subjective values of alternatives are discounted while not fixated. Finally, we performed a systematic search over a large model space, allowing us to evaluate the relative contribution of different forms of gaze-dependence and additional mechanisms previously not considered by gaze-dependent accumulation models. Gaze-dependence remained the most important mechanism, but participants with strong attraction effects employed an additional similarity-dependent inhibition mechanism found in other models of multi-alternative multi-attribute choice.

Introduction

Imagine you inherited money from a distant relative and you need to decide how to invest it. You reach out to your trusted investment advisor who swiftly sends you a brochure with different investment alternatives. After reading the brochure, you look at the overview table provided on the final page and deliberate between two alternatives that particularly interest you, visually inspecting them, one at a time, comparing expected returns, associated risks, fees, and other attributes. Previous work has established that the role of visual attention in decision making under risk exceeds mere information sampling[1–3]. Instead, as in other forms of preferential and perceptual decision making, visual fixations have a constructive role in the decision process, so that alternatives that are looked at for a longer time are generally more likely to be chosen[4–10].

This association between gaze and choice has been formalized in gaze-dependent evidence accumulation models[1,2,5,6,10–13]. The attentional Drift Diffusion Model (aDDM)[6], Gaze-weighted Linear Accumulator Model[12,13], and others that apply explicitly to risky choice[1] assume that decision makers accumulate evidence in favour of each alternative until evidence for one alternative reaches a threshold and a decision is made. Crucially, accumulation rates are assumed to depend on gaze allocation, so that evidence for an alternative is discounted while it is not fixated. Prior work has tested these assumptions in simple decisions under risk, involving two risky gambles with two equally probable outcomes[3] or two risky gambles described by outcome and probability[1]. Like the example of choosing an investment plan, however, many real-life decisions are more complex than simple binary choice. For example, investment decisions can involve more than two alternatives that vary on multiple attributes (such as expected return, associated risk, and fees), including both sets of options with similarly low expected returns (e.g., government bonds, fixed deposits) and sets of other riskier options promising larger gains at higher risk (e.g., stocks or derivatives).

Crucially, choices in these multi-alternative, multi-attribute settings pose a challenge to many traditional models of risky choice, including Expected Utility Theory (EU)[14] and Prospect Theory[15,16], which obey basic axioms of rational choice like independence of irrelevant alternatives (IIA)[17]. Briefly, IIA asserts that the preference between two alternatives should not depend on other alternatives. At least three context effects – the attraction[18], compromise[19], and similarity[20] effects – show that IIA is frequently violated: The attraction effect describes an increase in preference for an alternative following the addition of a similar, but slightly inferior alternative. The compromise effect describes an increase in preference for an alternative after a third alternative was added that makes it appear as intermediate. The similarity effect predicts that adding a third option that is similar to one of the original options, and equally appealing, will increase relative preference for the other, dissimilar alternative. While these effects are predominantly investigated using consumer goods, some studies also found them to affect choices between risky gambles[18,20–23].

Generally, models that assume that each alternative can independently be assigned a single scale value denoting its utility (fixed utility or simple scalability models[24]) cannot account for the observed violations. In contrast, models that can predict the effects often assume that preferences are constructed by means of comparisons between alternatives on different attribute dimensions[20,25–28], and employ additional psychological or neurobiologically-inspired mechanisms, such as loss-aversion or inhibition between alternatives.

While current gaze-dependent accumulation models of risky choice assume that each alternative can be assigned a singular underlying scale value (e.g., expected utilities for each alternative), choices are also assumed to be constructed in an accumulation process that remains malleable due to its dependence on gaze. Therefore, gaze-dependent accumulation models potentially could account for violations of IIA or regularity if the distribution of gaze itself changed depending on the decision context (e.g., shift towards dominant alternatives in the attraction effect, compromise alternatives in the attraction effect, or dissimilar alternatives in the similarity effect). Initial evidence for such an attention-shift in the attraction effect comes from a study of intertemporal choice[29]. Furthermore, recent work has suggested gaze-dependent evidence accumulation as a model of multi-alternative, multi-attribute choice in the absence of context effects[2,30].

Here, we test to what extent gaze-dependent accumulation can explain risky choices in the presence of context effects. We replicate the attraction and compromise effects in a task where participants made repeated choices in a multi-alternative, multi-attribute setting involving risky gambles. We compare a model of gaze-dependent accumulation adapted from previous work on binary risky choice[1] with an established model of context-dependent multi-alternative, multi-attribute choice without gaze-dependence and with other traditional models of risky choice. Notably, the gaze-dependent accumulation model did not include any dedicated mechanisms to produce context effects. Our task induced a set of context effects across the group, with substantial variability between individuals, providing a complex testing scenario to compare gaze-dependent accumulation models with competing theoretical accounts. We found the gaze-dependent model to give the overall best account of the choice data, while underestimating strong attraction effects of some individuals. A systematic search over a large model space combining features across model classes confirmed that all participants' behaviour was best described by a gaze-dependent accumulation process, but that individual differences in attraction effect strengths are predicted best by variants integrating mechanisms from other models.

Results

Choice task

In our experiment, 40 participants made repeated decisions between three all-or-nothing gambles, each described by a probability p to win an outcome m and nothing otherwise (Fig. 1a). We recorded

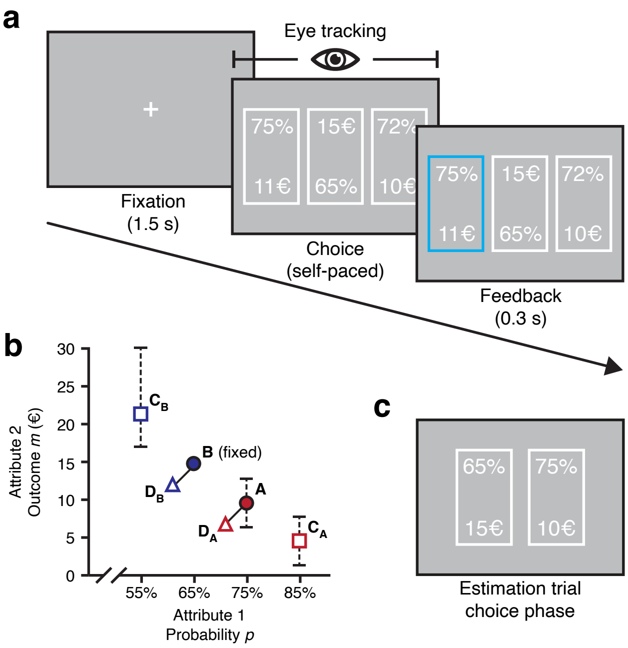


Figure 1. Choice task and attribute space of stimuli. **(a)** Choice task. Trials started with a central fixation cross for 1.5 s. Next, three gambles were presented. Participants made a choice by button press, without time limit. Eye movements were recorded during the choice phase. Finally, brief feedback on the choice (but not on the gamble outcome) was displayed. **(b)** Attribute space. Each gamble is described by two attributes: Probability p and outcome m. The core options A and B were presented in every trial, along with one of four decoy alternatives: Compromise decoys (C_A or C_B) and asymmetrically dominated decoys (D_A or D_B) are expected to elicit compromise and attraction effects, respectively. Dashed lines indicate possible stimulus placements. **(c)** Exact positions for C_B, A and C_A were calibrated for pairwise indifference in separate binary choice estimation blocks for each individual. Differences in outcome between neighbouring alternatives was not less than 2 EUR. Dominated decoys D_A and D_B were always 2% and 1 EUR worse than A and B respectively.

participants' eye movements during the task with an eye tracker (see Methods for details). Participants were instructed that after completing the task, their chosen gamble from one randomly determined trial would be played out for an additional bonus payment. In contrast to hypothetical choices between consumer goods described on attribute dimensions like quality and price, risky gamble stimuli offer a high amount of control over their attributes and a straightforward way to incentivise choices. The gamble stimuli were designed to elicit attraction and compromise effects and were individually tailored to account for each participant's risk preferences (Fig. 1b). Participants performed three pairs of indifference-estimation and experimental blocks, for a total of 225 experimental trials (see Methods for additional details on the experimental procedure and gamble stimuli). In indifference-estimation blocks, participants made repeated choices between pairs of gambles with different probabilities (Fig. 1c). Gamble outcomes m were adjusted according to participants' choices so that four approximately equally preferred gambles C_B, B, A, and C_A were constructed with winning probabilities p = 55%, 65%, 75%, and 85%. Additionally, two asymmetrically dominated decoys D_A and D_B were defined to be 2% and 1 EUR worse than gambles A and B, respectively. Following this indifference estimation, participants performed an experimental block of 75 ternary choice trials, 32 of which were compromise trials balanced with respect to the target option (i.e., 16 choice sets $\{A, B, C_A\}$, 16 choice sets $\{A, B, C_B\}$), 32 attraction trials (16 choice sets $\{A, B, D_A\}$, 16 choice sets $\{A, B, D_B\}$) and 11 distractor trials showing randomly created options with expected value of 10 EUR and low (5–33%), medium (34–64%) and high (65–95%) probability p. Uniform noise was added to all outcomes m (±0 EUR, +1 EUR) and probabilities p (–3%, ±0%, +3%) in each trial. Asymmetrically dominated decoys D_A and D_B received the same noise as the target option to preserve the dominance relation.

Context effects

We first analysed participants' choice behaviour to test whether their choices could be influenced by the set of available alternatives (see Table 1 and Fig. 2). Participants rarely chose the dominated decoys in attraction trials (mean ± s.d. = 0.02 ± 0.02 of attraction trials), indicating that participants understood the dominance relationships among the stimuli. In compromise trials, participants chose (non-dominated) decoy alternatives more frequently (mean ± s.d. = 0.27 ± 0.16). In particular, the decoy with the highest possible outcome C_B ($p \approx 55\%$, $m \approx 18$ EUR) was chosen most frequently when it was available. Note that decoy choices in compromise trials are expected, as extreme decoys were specifically calibrated to be approximately equally preferred to the core options.

We tested the presence of the attraction and compromise effects by first computing the relative choice

Trial type	Target	Competitor	Decoy
D_A	0.60	0.39	0.01
D_B	0.53	0.45	0.02
C_A	0.46	0.37	0.16
C_B	0.32	0.31	0.37

Table 1. Relative choice frequencies across participants in the four trial types. Across the group, target options were chosen more frequently than competitors in both types of attraction (D_A and D_B) and compromise trials (C_A, C_B). Dominated decoys were almost never chosen. In compromise trials including the high-outcome decoy C_B , the decoy was chosen more frequently than core options.

share of the target alternative (RST)[31] for each individual, and separately for attraction and compromise trials. The RST is computed as

$$RST = \frac{N_{\text{target}}}{N_{\text{target}} + N_{\text{competitor}}} \tag{1}$$

where N is the number of times an alternative was chosen. Given the balanced design, where each core alternative acts as a target and competitor equally often, the RST should be close to 0.5 if no context effects are present. If, however, the RST is different from 0.5, a systematic context effect is indicated. We tested whether the mean RST significantly differed from 0.5 across participants by computing its 95% highest posterior density interval (HDI_{95}) using Bayesian estimation (BEST)[32,33].

We find evidence for the attraction effect: The mean RST in attraction trials differed meaningfully from 0.5 (mean RST = 0.56, HDI_{95} = [0.51, 0.60], mean $d = 0.49$, HDI_{95} = [0.10, 0.80]). 25 of 40 (62.5%) participants had an RST above 0.5 in attraction trials. Notably, similar to previous work[34] a subgroup of participants (9 of 40, 22.5%) showed particularly strong attraction effects with individual RSTs above 0.7. We could not, however, find evidence that these individuals used the dominance relationship as a simplifying choice rule (see Supplementary Note 2). To allow comparisons with other studies that quantified context effects as the difference between choice shares between targets and competitors, we also report these differences: In attraction trials, the average difference was 0.12 (HDI_{95} = [0.01, 0.20], mean $d = 0.48$, HDI_{95} = [0.10, 0.85], Fig. 2b).

We only found weak evidence for the compromise effect using the gamble stimuli: The mean RST in compromise trials was 0.53, but its estimated HDI_{95} did not exclude 0.5 (HDI_{95} = [0.49, 0.57], 91.1% of posterior mass above 0.5, mean $d = 0.23$, HDI_{95} = [-0.11, 0.59]). 26 of 40 (65%) participants showed an RST above 0.5 in compromise trials. The mean difference between choice shares for targets and competitors was 0.05 (HDI_{95} = [-0.01, 0.11], mean $d = 0.29$, HDI_{95} = [-0.04, 0.62], 95.8 % of posterior mass above 0; Fig. 2e). These results are similar to the marginal effects obtained using perceptual stimuli in previous work[35,36].

Similar to other work[26,31], we found a positive correlation between the effects across participants, even though it was weaker in our task ($r = 0.24$, HDI_{95} = [-0.05, 0.51], 94.5% of posterior mass above 0).

Taken together, we successfully induced context effects within our participant sample, with non-trivial variability in the strength of those effects across individuals. These data provide a complex testing scenario to investigate gaze-bias effects in multi-alternative multi-attribute choice and to compare gaze-dependent accumulation models with competing theories.

Context effects are present in relative dwell times

Next, we tested whether participants' eye movements were affected by the set of available alternatives, similar to their choices. We therefore computed each alternative's relative dwell time in a trial by summing the duration of all fixations durations towards it, and normalizing to the sum of all fixations in the trial. Both patterns of context effects of choice behaviour were present in the relative dwell data: Target options received greater relative dwell time than competitors in attraction (mean difference = 0.014, HDI_{95} = [0.004, 0.024], mean $d = 0.53$, HDI_{95} = [0.15, 0.93], Fig. 2c) and compromise trials (mean difference =

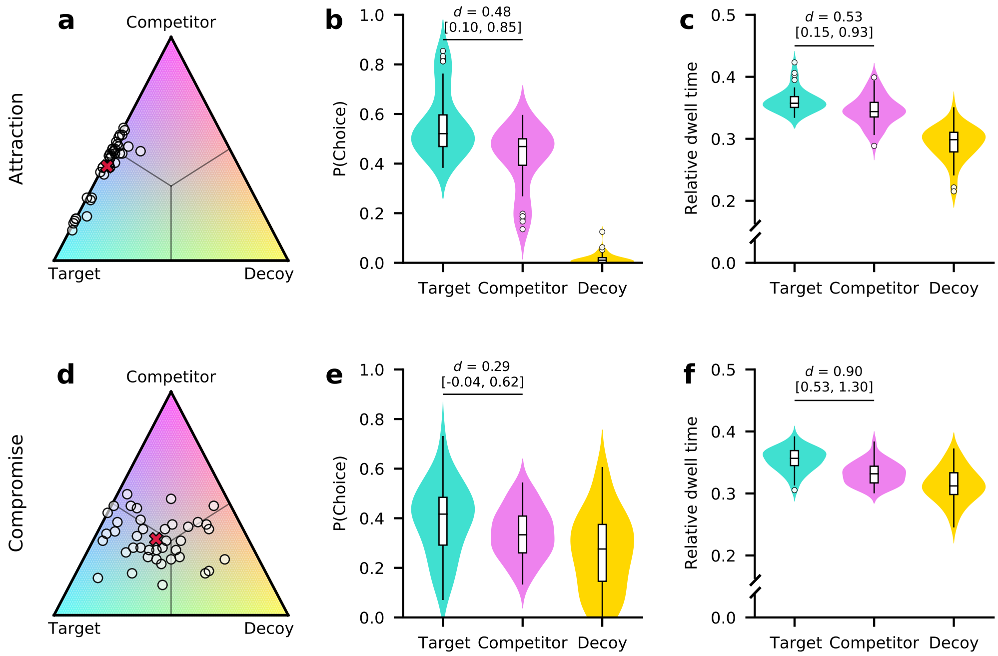


Figure 2. Context effects are present in choices and relative dwell time. Participants choices were influenced by asymmetrically dominated decoys and, to a lesser extent, by extreme compromise-making decoys. **(a, d)** Ternary plots of individual relative choice frequencies for target (lower left, teal), competitor (top, pink), and decoy (lower right, yellow) alternatives in attraction **(a)** and compromise **(d)** trials. Each dot represents one participant. The position on the simplex indicates relative choice frequencies for alternatives. Straight lines from the centre indicate equal frequencies for two alternatives. The red "x" indicates the group average. **(b, e)** Relative choice frequencies in attraction **(b)** and compromise **(e)** trials. In attraction trials, some participants strongly favoured the target alternative and almost no decoy choices were made. While target alternatives are still chosen more frequently than competitors in compromise trials, the effect is less pronounced, and extreme decoys are still chosen frequently. **(c, f)** Relative dwell time towards alternatives. In both, attraction **(c)** and compromise **(f)** trials, target alternatives received greater relative dwell times than competitors. d denotes Cohen's d from paired BEST analysis with HDI_{95} given in brackets. Violin plots show kernel density estimates of distributions of individual values. Box plots mark lower and upper quartiles and median. Whiskers extend from first and last datum within 1.5 times the interquartile range from lower and upper quartiles, respectively. Values outside this range are indicated by open circles.

0.022, HDI_{95} = [0.014, 0.030], mean $d = 0.90$, HDI_{95} = [0.53, 1.30], Fig. 2f). Targets and competitors also received greater relative dwell time than decoys in both trial types.

To better understand participants' eye movements over the course of the trial, we performed additional exploratory analyses of gaze location and transitions. We found an increasing association between gaze and choice over the trial, and longer gaze towards targets, even accounting for choice. Information search occurred both within and between alternatives, with slightly more transitions within alternatives. We refer to Supplementary Note 1 for additional details on these analyses.

Behavioural modeling and model comparison

Given that participants' choices and eye movements exhibited modulation by the context of available options, we next tested whether their behaviour could be described using a gaze-dependent accumulation model, and how such a model performs in comparison to extant theories of multi-alternative multi-attribute choice. To this end, we adapted a recently developed, gaze-dependent leaky accumulator model of two-alternative risky choice[1] to the three-alternative scenario. We refer to this model as the Gaze-dependent Leaky Accumulator (GLA). It assumes that subjective utilities for each alternative i are constructed as in Cumulative Prospect Theory[16] by first applying a probability weighting function, that transforms objective probabilities p into subjective decision weights. Outcomes m are transformed

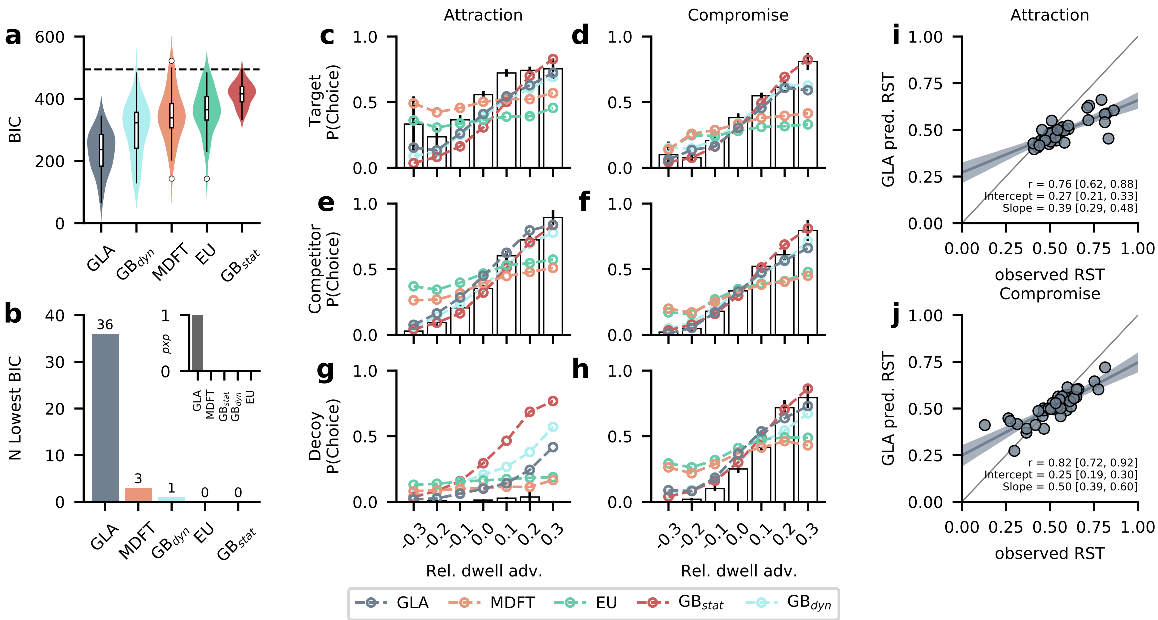


Figure 3. Model comparison and predictions. (a) The gaze-dependent leaky accumulator (GLA) provided the best fit (lowest mean BIC) across participants, followed by the dynamic gaze baseline model, MDFT, EU, and the static gaze baseline model. The dashed line indicates the BIC of the random choice baseline model. Violin plots show kernel density estimates of distributions of individual values. Box plots mark lower and upper quartiles and median. Whiskers extend from first and last datum within 1.5 times the interquartile range from lower and upper quartiles, respectively. Values outside this range are indicated by open circles. **(b)** The GLA fitted most (36 of 40, 90%) participants best, with a protected exceedance probability of 1 (inset). **(c-h)** Observed and model-predicted probability of choosing the target **(c, d)**, competitor **(e, f)**, or decoy **(g, h)** alternatives in attraction **(c, e, g)** and compromise trials **(d, f, h)** as a function of relative dwell time advantage. Relative dwell time advantage was computed as relative dwell time towards an alternative minus the mean relative dwell time to all other alternatives. White bars and error bars show mean ± s.e. observed data from even-numbered trials. Model predictions (coloured lines) are based on 50 simulations of each odd-numbered trial. **(i, j)** Observed and predicted RST of the best-fitting GLA for attraction **(i)** and compromise **(j)** trials. Each circle represents one participant. The winning model's predicted context effect sizes correlated significantly with the observed ones. Strong context effects, however, were underestimated, as indicated by the reduced slopes.

into utilities using a power function. Next, this model assumes that for each alternative i subjective utilities x_i^{GLA} are accumulated with leakage over the time course of each trial and that utilities are discounted while they are not fixated. At each fixation n, accumulators evolve according to

$$X_i^{GLA}(n) = \begin{cases} (1-\lambda)\cdot X_i^{GLA}(n-1) + 1\cdot x_i^{GLA} & \text{if } i \text{ fixated} \\ (1-\lambda)\cdot X_i^{GLA}(n-1) + \theta\cdot x_i^{GLA} & \text{otherwise} \end{cases} \tag{2}$$

where all $X_i(0) = 0$. Similar to the aDDM, the θ parameter $(0 \leq \theta \leq 1)$ controls discounting of unattended alternative utilities. The λ parameter $(0 \leq \lambda \leq 1)$ controls the strength of the accumulation leak. We computed choice probabilities from this model by applying the soft-max choice rule (Eq. (4)) over the final accumulator values X_i^{GLA}.

Additionally, we fitted an implementation of Multialternative Decision Field Theory (MDFT)[25], representing the class of dynamic multi-attribute, multi-alternative choice theories that have been designed to explain context effects. In contrast to GLA, MDFT assumes that preferences evolve dynamically over time by accumulation of attribute-wise comparisons between alternatives, and alternatives inhibit each other depending on their distance (or similarity). MDFT does not assume any influence of gaze on the decision process. We also included three baseline models in our comparison. First was a standard Expected Utility model (EU)[14] that did not consider gaze data. Second, a static gaze baseline model GB_{stat}

predicted choices only from trial-aggregated gaze towards each alternative, ignoring outcome values and outcome probabilities. Third, we constructed a dynamic gaze baseline model GB_{dyn} that, like GLA, assumed leaky evidence accumulation over fixations, but ignored attribute values, using only the sequence of fixations in each trial to predict choices. See Methods for details on these models' implementation and parameter estimation.

All models were fit individually to the choice data of each participant. We compared the models using the Bayesian Information Criterion (BIC)[37], which penalizes more complex models and therefore accounts for differences in complexity between models.

Across the group, all models performed better than the random baseline model (Fig. 3a). The GLA had the lowest BIC (mean $\pm$ s.d. $= 230.63 \pm 66.80$), followed by the dynamic gaze baseline model GB_{dyn} (mean $\pm$ s.d. $= 302.46 \pm 86.15$), MDFT (mean $\pm$ s.d. $= 345.77 \pm 82.44$), EU (mean $\pm$ s.d. $= 360.57 \pm 72.56$) and the static gaze baseline GB_{stat} (mean $\pm$ s.d. $= 414.81 \pm 36.81$). Simulating choices from the models using maximum-likelihood estimates, the proportion of correctly predicted choices was 74.2% for GLA, 62.8% for GB_{dyn}, 57.9% for MDFT, 53.8% for EU and 45.7% for GB_{stat} (see Supplementary Fig. 4 for distributions of model-predicted choice probabilities). Note, that target and competitor options (and decoys in compromise trials) were designed to be closely matched in value, resulting in trials with high choice difficulty, limiting overall model performance. On the individual level, based on lowest BIC scores, the majority (36 of 40, 90%) of participants were best described by GLA. Three participants (7.5%) were best described by MDFT, and one by the dynamic gaze-baseline model (Fig. 3b). Protected exceedance probabilities[38], which measure the likelihood that model is more frequent than all others, unambiguously identified GLA as the most likely model ($pXP_{GLA} = 1$, Fig. 3b inset).

The estimates for GLA's gaze-discount parameter θ, which describes the degree to which alternative's values are attenuated while not fixated, indicate that participants exhibited a moderate gaze-discount on average: θ estimates ranged from 0.13 (strong gaze-discount) to 0.95 (almost no gaze-discount), with mean $\pm$ s.d. $= 0.69 \pm 0.18$. Estimates for the leak parameter ranged from 0.08 (almost no leak) to 0.65 (moderately strong leak), with mean $\pm$ s.d. $= 0.29 \pm 0.20$. Individual parameter estimates of all GLA parameters are plotted in Supplemental Fig. 3 and summarised in Supplemental Table 1.

Given that MDFT outperforms utility-based models when choices are influenced by context[31], we tested whether model fit of MDFT (relative to GLA) was higher for participants with higher RST. In other words: Does MDFT perform better for stronger context effects? Overall, 7 out of 9 participants with an RST above 0.7 in attraction trials were clearly best described by GLA. While we found that the relative superiority of GLA over MDFT decreased with strong attraction effects, this was not the case for compromise trials (Supplemental Fig. 5), suggesting that some features of MDFT might help capture strong attraction effects.

As an additional indicator of model fit, we tested whether the models could quantitatively reproduce the observed positive association of gaze and choice (Supplemental Fig.s 1d, j, 2). Specifically, following previous work[6,11], we inspected the model-predicted probability of choosing an alternative as a function of its relative dwell time advantage: the difference in relative dwell time towards an alternative minus the mean relative dwell time to other alternatives. The probability of choosing an alternative generally increased with its relative dwell time advantage (Fig. 3c-h), except for dominated decoys in attraction trials, which were not chosen even if looked at longer than other alternatives (Fig. 3g).

All gaze-dependent models were able to capture this positive association. Note, however, that they also predicted choices of dominated decoys in attraction trials, if decoys had a large dwell time advantage (Fig. 3g). In this case, GLA performed better than GB_{dyn} and GB_{stat}, as it predicted fewer decoy choices. However, MDFT and EU generally could not capture the empirical association of gaze and choice; they predicted too many choices of alternatives with negative dwell time advantage, and too few choices of alternatives with positive dwell time advantage. Nor did they predict choices of dominated decoys, even if the decoy was looked at longer.

Finally, we investigated the ability of the best-performing model to predict individual differences in context effect strengths. Therefore, we predicted choices from the fitted GLA model and correlated the resulting RST with the observed data. Predicted RST correlated significantly with observed ones in attraction ($r = 0.76$, $HDI_{95} = [0.62, 0.90]$, Fig. 3i) and compromise ($r = 0.82$, $HDI_{95} = [0.72, 0.92]$, Fig. 3j) trials. However, the model underestimates large deviations from RST = 0.5, suggesting that its gaze-discount mechanism can capture the qualitative pattern of context effects but not their expression in participants with extreme RSTs.

Inspection of predicted choice probabilities (Supplemental Fig. 4) shows that, on average, GLA predicted high probabilities for the empirically chosen alternative (indicating good overall fit), but comparable proportions of target and competitor choices (resulting in reduced RST). Other models, like MDFT, predicted larger differences between target and competitor choices for some participants, but assigned lower probability to empirically chosen alternatives (resulting in overall inferior fit).

Systematic exploration of a large model space

The model comparison identified the advantage of the gaze-dependent accumulation model over its competitors. To better understand the contribution of individual model mechanisms (such as accumulation leak, inhibition between alternatives, or the gaze-dependent discount) to model performance, we performed a search across a large, systematically designed model space, in a so-called "switchboard analysis"[39]. Here, the idea is to isolate, group and exhaustively combine mechanistic assumptions of different models. Each group of mechanisms is considered a switch that can take different levels (e.g., an "inhibition" switch could take the levels "distance-dependent" as in MDFT, "constant" or "none"). Ultimately, this approach can be used to gauge the contribution of individual model mechanisms (opposed to evaluating whole models or model classes as in the more traditional model comparison presented above). In addition, it provides a systematic way to generate novel hybrid models, combining mechanisms from different *a priori* defined models.

This analysis approach further allowed us to test different assumptions about the functional forms of the gaze bias mechanism (e.g., as discounting non-fixated alternatives' values, controlling accumulation leak, among others[40]). We therefore expanded the range of gaze-dependent mechanisms from the original set of models to include additional eye-movement related switches, like attribute- and alternative-wise gaze-discounts, gaze-dependent leakage and inhibition. This allowed us to test if gaze-bias implementations different from the ones usually used in gaze-dependent accumulation[3,6,10–13,41,42].

Our switchboard comprised a total of 192 model variants, derived from six switches that were combined exhaustively: Attribute integration (additive vs. multiplicative), evidence comparison (independent accumulation for each alternative or comparative accumulation of evidence relative to other alternatives), alternative-wise and attribute-wise gaze-discount (included or not), accumulation leak (none, constant or gaze-dependent) and inhibition between alternatives (none, constant, gaze-dependent or distance-dependent). The models generally resembled the form of the *a priori* defined GLA, but with substantial differences depending on the specific set of switch levels (Fig. 4a; see Methods and Supplemental Table 2 for details on the framework and switch levels). As before, each model variant was fit to the individual data of each participant and model performance was evaluated based on the BIC score. Note that the GLA coincides with one of the 192 variants (variant A in Table 2; multiplicative attribute integration, alternative-wise gaze-discount, constant leak, no inhibition, no comparison). Similarly, one variant (not in Table 2) conceptually resembles MDFT in some, but not all aspects (additive attribute integration, comparative evidence accumulation, constant leak, distance-based inhibition, strong attribute-wise gaze-discount).

Best model mechanisms

On the level of model mechanisms, multiplicative attribute integration outperformed additive integration (mean $\mathrm{BIC}_{\mathrm{mul.}} = 312.25$, mean $\mathrm{BIC}_{\mathrm{add.}} = 333.20$; Fig. 4b), inclusion of an alternative-wise gaze-discount (mean $\mathrm{BIC}_{\mathrm{GD\ alt.}} = 289.13$, mean $\mathrm{BIC}_{\mathrm{no\ GD\ alt.}} = 356.32$), but not attribute-wise gaze-discount yielded lower BIC scores (mean $\mathrm{BIC}_{\mathrm{GD\ att.}} = 324.11$, mean $\mathrm{BIC}_{\mathrm{no\ GD\ att.}} = 321.33$). Other gaze-dependent mechanisms also improved model fit: Variants with gaze-dependent leak yielded lower BIC scores than variants with constant or no leak ($\mathrm{BIC}_{\mathrm{gaze}} = 292.06$, $\mathrm{BIC}_{\mathrm{constant}} = 304.84$, $\mathrm{BIC}_{\mathrm{none}} = 381.49$). Gaze-dependent inhibition performed better than constant, none or distance-dependent inhibition ($\mathrm{BIC}_{\mathrm{gaze}} = 305.64$, $\mathrm{BIC}_{\mathrm{constant}} = 316.06$, $\mathrm{BIC}_{\mathrm{none}} = 311.75$, $\mathrm{BIC}_{\mathrm{distance}} = 351.57$). In summary, all mechanisms that allowed a model to exploit the positive association between gaze and choice improved model fit on average.

Counting switch-values of individually best fitting variants, most participants were best described by model variants with multiplicative integration, with alternative-wise and no attribute-wise gaze-discount, with constant accumulation leak parameter and no inhibition (Supplemental Fig. 6).

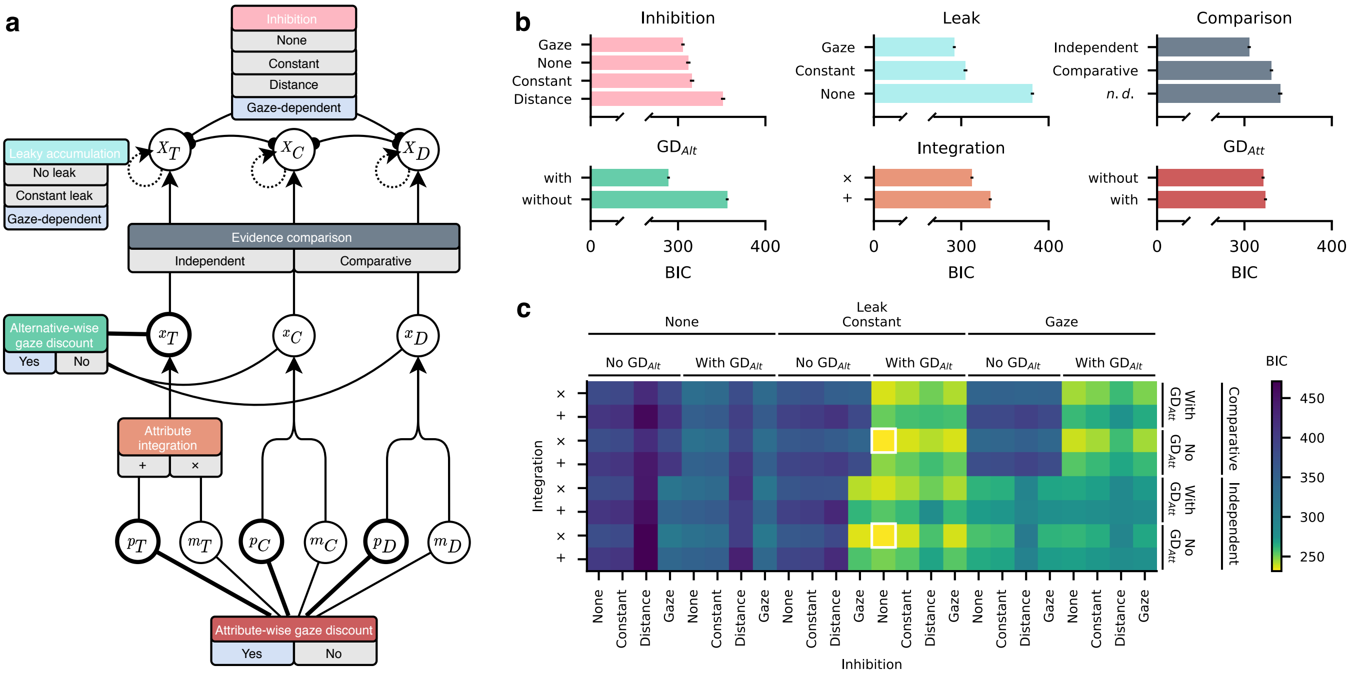


Figure 4. Switchboard analysis. (a) Overview of general switchboard framework and individual switches from which individual variants are constructed by setting the switches to a set of levels. Gaze-dependent switch levels are shaded blue. Attributes can be discounted based on gaze (lower level) and are integrated into alternative values (middle level). Alternative values can be discounted based on gaze, compared, and integrated (upper level) with leak and inhibition. **(b)** Average model fit associated with each switch's levels. Each bar shows the average BIC for all model variants that had the respective switch set to this level (e.g., first panel, top bar: average BIC of all variants with gaze-dependent inhibition). Gaze-dependent inhibition and leak, independent evidence accumulation, alternative-wise gaze-discount, multiplicative attribute integration, and no attribute-wise gaze-discount yielded lower BIC on average. **(c)** Overview of mean BIC for each of 192 model variants. More yellow colours indicate lower BIC and better model fit. The variant with the lowest BIC is identical to the GLA (alternative-wise, no attribute-wise gaze-discount, multiplicative attribute integration, constant leak, and no inhibition) and is outlined in white. Note that some variants were mathematically equivalent (see main text and Methods) including the variant with lowest BIC, which is therefore highlighted twice.

Best model variants

The overall best-fitting model variant was the variant identical to the GLA (Fig. 4c, Supplemental Table 3): It included multiplicative attribute integration, an alternative-wise gaze-discount, and constant leak. Note that our analysis did not allow us to distinguish independent and comparative accumulation for this variant, because they yield equivalent results. All of the ten best performing models used multiplicative attribute integration, and most used an alternative-wise but no attribute-wise gaze-discount and constant leak (Supplemental Table 3). Results were more ambiguous for the comparison and inhibition mechanisms.

Notably, one of the best ten variants employed a distance-dependent inhibition mechanism, and used other features resembling MDFT like constant leak, and accumulation of comparative values. Unlike MDFT, but like GLA, this variant used an alternative-wise gaze-discount, no attribute-wise mechanism of attention, and multiplicative integration of attributes. While not achieving the best overall fit, this hybrid variant performed significantly better than the original MDFT implementation (mean $\text{BIC}_{\text{MDFT}} = 345.77$; Fig. 3a).

The GLA variant also described 17 of 40 (42.5%) participants best (Table 2). Another similar variant with additive attribute integration described an additional four participants best. An additive relationship between attributes is typically assumed by models of multi-attribute multi-alternative choice[25,27]. Furthermore, additive integration of probability and outcome has recently been suggested as an alternative to multiplicative mechanisms and has been demonstrated to be equivalent for particular parameterisa-

	N	$\mathbf{GD_{Alt}}$	$\mathbf{GD_{Att}}$	**Leak**	**Inhibition**	**Integration**	**Comparison**
A	17	Yes	No	Constant	None	Multiplicative	*n.d.*
B	9	Yes	No	Constant	Distance	Multiplicative	Comparative
C	4	Yes	No	Constant	None	Additive	*n.d.*
D	3	No	No	Gaze	None	Multiplicative	Independent
E	2	No	No	Constant	Gaze	Multiplicative	Independent
F	1	Yes	No	Gaze	Distance	Multiplicative	Comparative
G	1	Yes	No	Gaze	Constant	Multiplicative	Comparative
H	1	Yes	Yes	Constant	Distance	Multiplicative	Independent
I	1	Yes	No	Constant	Gaze	Multiplicative	Comparative
J	1	Yes	No	Gaze	None	Multiplicative	Comparative

Table 2. Overview of individually best fitting model variants. N indicates the number of participants best described by the variant described in the row. The top variant (A) coincided with the GLA model. Note that all individually best fitting models had some form of gaze-dependence (blue shaded cells, mostly alternative-wise gaze-discount). *"n.d."* denotes variants where comparison mechanisms were not distinguishable by the analysis.

tions and parameter values[43]. Five participants were best described by variants similar to the GLA, but using gaze-dependent leakage or inhibition instead of an alternative-wise gaze-discount. Note that gaze-dependent inhibition and leakage mechanisms can act similarly to an alternative-wise gaze-discount: All three mechanisms effectively reduce accumulated evidence of non-fixated alternatives. While the alternative-wise gaze bias mechanism applies a constant discount to the accumulation inputs, the gaze-dependent inhibition mechanism is proportional to the accumulated evidence of the currently fixated alternative, and applies to the already accumulated evidence of other options, not the inputs to the accumulation process. Gaze-dependent leakage similarly reduces already accumulated evidence, proportional to the momentary accumulator value.

Notably, nine participants (22.5%) were best described by the previously described hybrid variant using a distance-dependent inhibition mechanism (Table 2). Additional two participants were best described by other variants using distance-dependent inhibition in conjunction with an alternative-wise gaze-discount.

Hybrid variant

Finally, we analysed the hybrid model variant in more detail (variant B in Table 2), which described 9 participants best. This variant performed especially well for participants with large attraction effects (Fig. 5a), whereas GLA best described the majority of participants with attraction RST around 0.5. In contrast, better-performing variants could not be separated by compromise effect strength (Fig. 5b). The hybrid variant correctly predicted individual differences in the attraction effect (correlation between observed and predicted RST: $r = 0.92$, $\mathrm{HDI}_{95} = [0.86\ 0.96]$; Fig. 5c). Here, it performed better than the GLA model (Fig. 3i), as it was also able to capture the behaviour of participants with particularly strong attraction effects: Using distance-dependent inhibition, it was able to predict high choice probabilities for target alternatives in attraction trials for some participants, and fewer choices of competitor and dominated decoy alternatives (Supplementary Fig. 4f). Predictions of individual RST in compromise trials were almost identical between the two models ($r = 0.81$, $\mathrm{HDI}_{95} = [0.70, 0.91]$; Fig. 5d, see Fig. 3j for GLA).

The hybrid variant used an alternative-wise gaze-discount and could thus accurately capture the relationship between relative dwell advantage and choice (Fig. 5e, f). Again, it predicted an overall higher probability of target choices than GLA (Fig. 5e), and this was primarily driven by the hybrid variant's predictions for individuals with strong attraction effects (Supplemental Fig. 7). There was no meaningful difference between the two models in compromise trials (Fig. 5f).

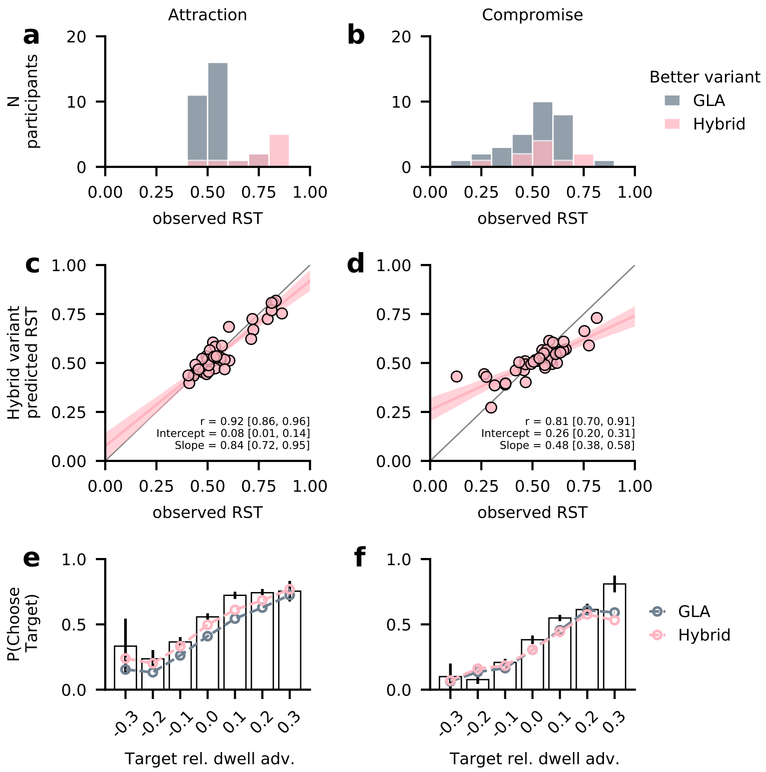


Figure 5. Hybrid model variant details. (a, b) Number of participants better described by the hybrid variant (pink) or the GLA (grey), dependent on strength of attraction **(a)** and compromise **(b)** effects. Participants with strong attraction effects were better described by the hybrid variant. **(c, d)** Individual observed and predicted RST in attraction **(c)** and compromise trials **(d)**. Compared to GLA (Fig. 3i, j), the hybrid model better predicted strong attraction effects for some participants. Predictions of compromise effects are similar. **(e, f)** Observed and model-predicted probability of choosing the target alternative, depending on the target's relative dwell time advantage. Like other gaze-dependent models (Fig. 3), the hybrid variant generally captured the positive association between gaze and choice. In contrast to GLA, however, it predicted an overall higher probability of choosing the target in attraction trials **(e)**. Predictions in compromise trials **(f)** are similar to GLA. White bars and error bars indicate mean ± s.e. observed data from even-numbered trials. Model predictions are based on 50 simulations for each odd-numbered trial.

Discussion

In this study, we investigated whether risky choice behaviour could be characterized by a gaze-dependent evidence accumulation framework, especially when choices are influenced by the context of available alternatives. In line with previous findings, we found choice behaviour to be context-dependent[44], but also subject to large interindividual differences. Importantly, participants' gaze behaviour was also modulated by the context of available alternatives, allowing a simple gaze-dependent evidence accumulation model derived from prior work on binary risky choice[1] to provide the best description of their choices. Finally, in a systematic search across a large space of possible model variants, we showed that gaze-dependent accumulation describes all participants' behaviour best. Predicting data from participants with particularly strong attraction effects, however, required inclusion of an additional similarity-based inhibition mechanism.

Prior work could already demonstrate that choices between risky prospects can be influenced by the set of available alternatives, producing attraction[18,21–23], similarity[20], and other decoy effects[22]. These findings show that risky prospects, described by their winning probability and outcome, are subject to the same context-dependent influences as other multi-attribute stimuli, even though the natural (or normative) relationship between their attributes is multiplicative and not additive.

Our findings add to this literature, replicating the attraction effect and providing novel evidence for the compromise effect in risky choice. Notably, we find substantial individual differences in the strength of the attraction effect, ranging from almost no to large effects. The observed pattern includes a subgroup of participants predominantly choosing the target in attraction trials, that is also present in previous work in inference[34] and risky choice tasks[21]. Importantly, we could not find any evidence that these participants used simplifying choice rules based on the dominance relationship. The observed variability further emphasizes the importance of focusing on individuals' behaviour instead of group aggregates, if the goal is to understand the processes underlying individuals' choices[45].

Our study involved choices within sets of three risky gambles designed to elicit context effects. Prior research makes contradicting predictions about the direction of information search in this scenario: In the context of risky choice, empirical studies find a tendency towards within-alternative processing, disagreeing with non-compensatory, heuristic approaches and providing better support for compensatory strategies that assume integration of outcomes and probabilities[46–48] (but see refs.[49,50], or in the related domain of intertemporal choice: ref.[51]). In the domain of context effects, however, Noguchi & Stewart found pairwise comparisons between alternatives on single attribute dimensions to be the dominant pattern of information search[52]. They argue that these comparisons should form the basis of choice models describing context effects. Similarly, Marini *et al.* found that adding an asymmetrically dominated decoy to a choice set shifts eye movements towards the target's dominant attribute, and results in more transitions between target and decoy[29]. Cataldo & Cohen showed that the way information is displayed can influence the size and direction of context effects[53]: Alternative-wise presentation yielded similarity effects, whereas attribute-wise presentation, thought to induce comparisons between alternatives on single attributes, produced attraction and compromise effects. In line with the risky choice and context effects literature, we found participants to shift their gaze both within and between alternatives. While this does not constitute strong evidence for any particular process, this finding is compatible with current models of gaze-dependent accumulation in risky choice[1,3] and the GLA that assume within-alternative integration of probabilities and outcomes, and gaze-dependent accumulation and comparison processes to reach a decision.

Across decision making domains longer gaze towards an alternative is generally associated with a higher probability of choosing it[3–6,8,10,11,40,54–57]. This association is also present in choices between risky prospects[1,3,46,58]. While these results are mostly correlational, multiple studies found that manipulation of gaze towards an alternative increases its likelihood of being chosen, suggesting that gaze allocation influences choice[4,7–10,59].

The positive association between gaze and choice is also present in our data: Chosen gambles were looked at longer than others, and the effect increased over the course of a trial (see Supplementary Information). In addition, the probability of choosing an alternative increased with increasing relative dwell time. Gaze-dependent accumulation provides a formal account of the association between choice and gaze data, as unattended alternatives' accumulation is diminished, making them less likely to be chosen. Conversely, if context effects were present in participants' gaze, this would enable such models to predict context-dependent choice. Our data illustrate that this contextual modulation of gaze is indeed present.

Note that, in principle, GLA could produce a choice bias towards the alternative fixated last, by combining strong leakage with a strong gaze-discount: With a strong leak, predicted choices are influenced mainly by the information acquired in the final fixation. A strong gaze-discount could then bias choice towards the fixated alternative. The obtained parameter estimates, however suggest only moderate strengths of leak and discounting, indicating that the model's good performance was not purely driven by effects of last fixations, which are often directed to the chosen alternative.

Our results are closely related to recent work showing that another behavioural effect in multi-alternative, multi-attribute choice is mediated by visual attention: Addition of a third alternative to a choice set has been shown to affect choice accuracy through value-based attentional capture in choices between risky prospects[2] and food items[55]. This mediation through gaze, formalized by a gaze-dependent accumulation model, provided a better description of the observed data than competing accounts. Adding to other work implicating mechanisms of visual attention in the emergence of context effects[53,60], our work shows how gaze can mediate context effects in a similar way: Choice sets affect the distribution of gaze, which in turn affects the choice process.

Many traditional models of risky decision making assume that one scale value is assigned to each alternative independent of the presence of others, and that choice probabilities are directly derived from these values (e.g., Luce, 1959). These "simple scalability" theories include the most influential models of risky decision making (e.g., Expected Utility Theory[14]; and Prospect Theory,[15]). They obey rational axioms of choice like IIA and therefore cannot account for context effects by design[24]. To explain context effects, multiple competing accounts have been proposed[22,25–27,61–64] (see ref.[39], for a taxonomy of mechanisms and overview).

These models often assume that an alternative's value is computed in comparisons to other alternatives on single attributes[25,27], that the considered attribute dimension switches stochastically from moment to moment[25,27,61], and that choices result from accumulation (often imperfect, i.e., leaky) of evidence until a threshold is reached[25–27,61]. Switching between attribute dimensions can introduce correlations between accumulators for similar alternatives, generating similarity effects[25,27,39,61]. In order to account for other context effects, these models employ additional mechanisms: For example, loss aversion, that is, differential weighting of advantageous and disadvantageous comparisons can produce attraction and compromise effects[27]. Distance-dependent inhibition between alternatives can yield similar results, by inhibiting similar alternatives more strongly and bolstering alternatives that are similar but dominant[25]. Non-linear value functions discounting alternatives with extreme attribute values can produce compromise effects[26].

However, while they propose precise psychological processes leading up to decisions, their relationship to observable process data, like eye movement recordings, remains implicit. For example, the switching between attribute dimensions is often considered an attentional mechanism[25,27,39,61], yet it is assumed to occur at every time-step (e.g., millisecond), and therefore cannot be mapped to observable eye-movement data without additional assumptions. Notably, thus far models of context-dependent choice do not include any gaze-dependency in the decision process. This is in contrast to gaze-dependent accumulation models[1,5,6,11–13], which propose a formal account of the association between gaze and choice.

In our study, context-dependent choices were best explained by a straightforward three-alternative extension of a gaze-dependent accumulation model that was previously applied to binary risky choices[1]. This model assumes that each alternative can be assigned a value by multiplicative integration of probability and outcome attributes, independently of other alternatives. Unlike simple scalable theories, however, it accumulates these values in a gaze-dependent fashion until a choice is made. Through its gaze-dependence, this model was able to predict individual differences in context effects. It performed best even across a large space of models, which included variants using additive attribute integration, attribute-wise gaze-discount and accumulation of comparative values. Such variants resemble extant models of context-dependent choice (e.g., MDFT[25]), as they accumulate results from single attribute comparisons. Yet they performed worse, even when they included gaze-dependency. Our results thus question whether models of context-dependent choice must use attribute-wise comparisons over alternative-wise integration of attributes. However, we also found that strong context effects could be predicted best using an additional inhibition mechanism based on alternatives' similarity (which is comparative in nature), while still using alternative-wise valuation at its core, suggesting multiple parallel processes (i.e., within-alternative valuation *and* comparative mechanisms).

More generally, our results suggest that extant models of context-dependent choice are likely to benefit from implementing gaze-dependence. Even further, explicitly formalizing the relationship between model variables and eye movements yields testable predictions that can help distinguish and evaluate competing theories about the role of attention in context-dependent risky choice. The identified class of models is compatible with observed transition data, quantitatively captures the association of dwell time and choice probability, and uses the contextual modulation of gaze in addition to a distance-dependent inhibition mechanism to predict context effects.

Methods

Participants

We recruited 44 participants for the experiment. All participants were required to have normal or corrected to normal vision with soft contact lenses. Participants relying on glasses or hard contact lenses were excluded from participation to ensure good eye tracking quality. Four participants were excluded from the analyses: one due to a computer crash, two due to eye tracking calibration worse than 1.0° of visual angle, and one because of misunderstanding task instructions. The remaining 40 participants (25 female, 15 male) had mean ± s.d. age 27.2 ± 4.7. All participants received a base compensation of 8 Euros per hour and could win an additional bonus based on their choices during the experiment (see below). Written informed consent was obtained from all participants prior to the experiment. The experimental procedures were approved by Freie Universität's ethics committee.

Task and stimuli

Participants performed a value-based choice task with stimuli designed to elicit attraction and compromise effects (Fig. 1). Each trial started with a 1.5 s fixation cross at the screen centre. Then, three all-or-nothing gambles were presented next to each other on the screen. Gambles were described by a probability p to win outcome m (and winning nothing otherwise). Each gamble was enclosed by a rectangle. Gamble attributes p and m were arranged so that the vertical distance between two attributes of one option was equal to the horizontal distance between the centres of neighbouring alternatives. This distance was set to approximately 10.0° of visual angle. Alternative positions and attribute positions within each gamble were random in each trial. Participants were instructed to indicate their preference for one of the three gambles using their right hand and the keyboard's arrow keys. There was no time limit. After their choice, participants received a brief (0.3 s) feedback about their choice (but not about a gamble outcome).

Participants were instructed that after completing the task, one of the trials would be chosen randomly and the gamble chosen in this task would be played out for real money with a virtual wheel of fortune, using a later to be disclosed payment multiplier. This multiplier was set at 0.5 to scale winning bonuses to Freie Universität's payment standards.

Participants first performed three pairs of estimation and experiment blocks. Estimation consisted of a maximum of 30 trials with choices between two alternatives. These blocks served the purpose of determining individual indifference points for stimuli with varying levels of winning probability p in an adaptive and integrated fashion. Participants were asked to indicate their preference between a fixed reference gamble B ($p_B = 65\%$, $m_B = 15$ EUR) and less risky option A ($p_A = 75\%$, $m_A = 10$ EUR). The outcome m_A for option A is then either increased (if B was preferred) or decreased (if A was preferred) and the procedure repeated. Indifference points for options C_A and C_B with probabilities $p_{C_A} = 85\%$ and $p_{C_B} = 55\%$ were determined interleaved and in the same fashion. A single estimation block yielded a stimulus set with four options A, B, C_A and C_B designed in a way that option pairs $A - B$, $A - C_A$ and $B - C_B$ were approximately equally preferred and their distance in outcome mi was not less than 2 EUR. Additionally, asymmetrically dominated range-frequency decoy options D_A and D_B were introduced and designed to be 2% and 1 EUR worse than options A and B, respectively.

The following experimental blocks had 75 ternary choice trials each, 32 of which were compromise trials balanced with respect to the target option (i.e., 16 choice sets $\{A, B, C_A\}$, 16 choice sets $\{A, B, C_B\}$), 32 attraction trials (16 choice sets $\{A, B, D_A\}$, 16 choice sets $\{A, B, D_B\}$) and 11 distractor trials showing randomly created options with expected value of 10 EUR and low (5–33%), medium (34–64%) and high (65–95%) probability p. For each trial, uniform noise was added to each option's outcome m (±0 EUR, +1 EUR) and probability p (–3%, ±0%, +3%). Asymmetrically dominated decoys D_A and D_B received the same noise as their focal option, to preserve dominance relation.

Participants performed 25 practice trials not relevant for their payout under supervision of the experimenter.

Eye tracking

Participants' eye movements were recorded at 60 Hz using a table-mounted SMI Red eye tracker (Senso-Motoric Instruments, Teltow, Germany). Participants were placed approximately 60 cm in front of the screen and instructed to minimize head movements during the task. Before each block, the eye tracker was calibrated using a 5-point calibration and validation procedure until a spatial resolution smaller than 1.0° visual angle was achieved horizontally and vertically. Participants were instructed to re-centre their gaze on the central fixation cross between trials.

Eye tracking data was pre-processed according to the following procedures: First, fixations, saccades and blinks were detected using SMI's Event-Detector software. Minimum fixation duration for detection was left at the default setting (80 ms). Blinks and saccades were discarded. Fixations were truncated when participants made a keyboard response. Next, rectangular areas of interest (AOIs) were constructed around the six screen locations that displayed stimulus attributes. Fixations towards non-AOI regions of the screen were discarded if they were preceded and followed by fixations to different AOIs. If they were preceded and followed by fixations towards the same AOI, the non-AOI fixation was re-coded to that AOI, too[6,11]. Finally, the total dwell time towards each alternative and attribute in each trial was computed by summing all fixation durations towards respective AOIs. Relative dwell time was computed by normalisation to the sum of all dwells in the trial.

Behavioural modelling

Baseline: Random choice

The random choice model predicts equal choice probabilities $p = \frac{1}{3}$ for all three alternatives and serves as a benchmark against which other models can be compared.

Expected Utility

Expected Utility Theory (EU)[14] assumes that choice behaviour can be described as maximization of expected subjective utility. We computed subjective utilities of option outcomes m_i using a power function with one free parameter α:

$$U(m_i) = m_i^{\alpha} \tag{3}$$

Predicted choice probabilities were computed using a soft-max choice rule[65] over expected utilities $x_i^{EU} = p_i \cdot U(m_i)$:

$$p(x_i) = \frac{e^{\beta x_i}}{\sum_{j \in J} e^{\beta x_j}} \tag{4}$$

Here, J is the set of all available alternatives. The inverse temperature parameter β controls the degree of randomness in the choice (choices become more deterministic with larger β).

Multialternative Decision Field Theory

Multialternative Decision Field Theory (MDFT)[25] is a dynamic connectionist model for multi-attribute, multi-alternative choice. MDFT can predict similarity, attraction and compromise effects. The core principle of MDFT is the calculation of valences $V(t)$ at each point in time t. Valences are determined as

$$V(t) = CMW(t) + \varepsilon(t) \tag{5}$$

where M is a matrix containing all alternatives' attributes. $W(t)$ is a stochastic vector indicating the momentary weight distribution between attributes, according to a weight parameter w. C is a contrast matrix, designed to perform attribute-wise contrasts between one option and the mean of other options' attributes. Finally, $\varepsilon(t)$ is a stochastic normally distributed noise component. Preferences $P(t)$ are generated integrating all valences $V(t)$ up to time t:

$$P(t) = SP(t-1) + V(t) \tag{6}$$

Critically, S is a square feedback matrix, thought to reflect the neurobiological mechanism of lateral inhibition between alternatives. The diagonal elements of S determine how much the current preference state is influenced by the previous one, controlled by the decay parameter φ_2. Off-diagonal elements represent the feedback connections between alternatives. MDFT assumes that the connection strength between two alternatives depends on their perceived distance D in attribute space, scaled by sensitivity parameter φ_1. Taken together, the feedback matrix S is given by

$$S = I - \varphi_2 \exp(-\varphi_1 D^2). \tag{7}$$

Here, I is the identity matrix. D is a matrix containing pairwise distances between alternatives. We used the distance function formalized in Hotaling *et al.*[66], where the distance between two alternatives is expressed in dominance- and indifference-directions and additionally scaled in dominance direction with an overweighting parameter w_d. The preference vector $P(t)$ is asymptotically normally distributed with mean ξ and covariance matrix Ω[67], from which choice probabilities are derived as in the general Thurstone model[68].

In total, this MDFT implementation includes five free parameters: The attribute weight w, the decay parameter φ_2 , the sensitivity parameter φ_1, the overweighting parameter w_d and the variance σ of the noise component $\varepsilon(t)$.

Typically, MDFT assumes an additive relationship between attributes. To accommodate for this, we log-transformed stimulus attributes probability p and outcome m[21]. Stimulus attributes were also rescaled to a range between 0 and 1 as in previous studies[31].

Gaze-based models

We included three models that use gaze data to predict choices: Two baseline models that ignore stimulus information and predict choices only based on gaze data, and one model adapted from previous work[1] that combines stimulus and gaze information in a leaky accumulation framework.

Static gaze baseline model The first gaze-based model predicts choices from participants' cumulated dwell times towards each alternative. It assumes that preference strength $x_i^{GB_{\text{stat}}}$ for an alternative increases when it is fixated, irrespective of the its attributes p_i and m_i:

$$x_i^{GB_{\text{stat}}} = d_i \tag{8}$$

where d_i is the total dwell time (in s) towards alternative i in a trial. Preference strengths $x_i^{GB_{\text{stat}}}$ are transformed into choice probabilities using the soft-max function (Eq. (4)).

Dynamic gaze baseline model The second gaze-based model uses the whole sequence of fixations in a trial to predict choices. It assumes that at each fixation, evidence in favour of the fixated alternative is accumulated, and that accumulated evidence is subject to leak. Formally,

$$X_i^{GB_{\text{dyn}}}(n) = \begin{cases} (1-\lambda)\cdot X_i^{GB_{\text{dyn}}}(n-1) + 1 & \text{if } i \text{ fixated} \\ (1-\lambda)\cdot X_i^{GB_{\text{dyn}}}(n-1) + 0 & \text{otherwise} \end{cases} \tag{9}$$

where all $X_i^{GB_{\text{dyn}}}(0) = 0$. The λ parameter ($0 \leq \lambda \leq 1$) controls the strength of the accumulation leak. Choice probabilities are computed from the final accumulator values using the soft-max function (Eq. (4)).

Gaze-biased leaky accumulator model (GLA) Gaze-biased leaky accumulator model (GLA). Finally, following a recent study on binary risky choice[1], we adapted a leaky accumulator model[69], where option values are discounted depending on eye movements as in the aDDM[6,11] and the related GLAM[12,13].

Here, the subjective utility for each alternative is computed by first applying a probability weighting function[16], that transforms objective probabilities into subjective decision weights:

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}} \tag{10}$$

where γ ($0.28 \leq \gamma \leq 1$) is a free parameter controlling the shape of the weighting function. If $\gamma = 1$, the subjective weights equal the objective probabilities[70]. Outcome utilities are assumed to obtained from a power function as in the EU model (Eq. (3)). Subjective expected utilities are then given by $x_i^{GLA} = w(p_i) \cdot U(m_i)$.

Next, this model assumes that for each alternative subjective expected utilities are accumulated with leak and gaze bias over the time course of each trial. At each fixation n, accumulators evolve according to

$$X_i^{GLA}(n) = \begin{cases} (1-\lambda) \cdot X_i(n-1) + 1 \cdot x_i^{GLA} & \text{if } i \text{ fixated} \\ (1-\lambda) \cdot X_i(n-1) + \theta \cdot x_i^{GLA} & \text{otherwise} \end{cases} \tag{11}$$

where all $X_i(0) = 0$. The θ parameter ($0 \leq \theta \leq 1$) controls discounting of unattended alternative values. The λ parameter ($0 \leq \lambda \leq 1$) controls the strength of the accumulation leak.

Predicted choice probabilities are again computed from the soft-max function (Eq. (4)) over the final accumulator values X_i^{GLA}.

Parameter estimation

All models' parameters were estimated by minimizing the negative summed log-likelihood $-\ln(\hat{L})$ of observed choices under the model. Minimization was performed by a differential evolution algorithm[71] implemented in the SciPy Python library[72]. The algorithm was provided sensible *a priori* defined bounds for each parameter and initialized randomly.

Model comparison

Individually best-fitting models were identified based on the Bayesian Information Criterion (BIC)[37], computed for each model m as

$$BIC_m = -2\ln(\hat{L}) + \ln(n_{trials})k_m \tag{12}$$

where k_m is the number of free parameters of model m and $\ln \hat{L}$ is the summed log-likelihood of observed choices under model m.

Switchboard analysis

We performed a switchboard analysis, similar to the one performed by Turner *et al.*[39] to further investigate which components of the cognitive model are particularly important in predicting the data. In a switchboard analysis, a cognitive model of the decision process is built, where individual model mechanisms can take different forms, or levels, which can be switched and combined with each other. One switch, or node, could for example be the integration of attributes to form item values. This integration could happen multiplicatively, so that expected outcomes are computed by multiplying outcome value and probability. It could also occur in a weighted additive fashion, so that both outcome value and probability make independent contributions to overall item value[43]. In the switchboard analysis, model variants using both implementations, and combinations with all other levels of other nodes, are constructed and fit to the behavioural data.

The switchboard analysis included different eye-movement related nodes, such as attribute and alternative wise gaze biases or gaze-dependent leakage and inhibition, so that the mechanisms that describe the data best can be identified and their relative contribution to model fit can be measured. All switchboard models resembled the general form of the gaze-dependent accumulation model presented in Glickman *et al.*[1] and the GLA adaptation to three items (see above and Fig. 4a for a schematic). Here, evidence X_i in favour of each item is accumulated over individual fixations. Accumulation can be subject to gaze-discount effects (so that non-fixated items accumulate less evidence), leak and inhibition over time. Choice probabilities are computed using a soft-max function (Eq. (4)) over the final accumulator values. The general accumulation framework (in vector form, parallel for each item) is then given by

$$X(t) = S \times X(t-1) + \Theta C x \tag{13}$$

where S is a square feedback matrix, combining the effects of accumulation decay (on its diagonal elements) and inhibition between accumulators (on off diagonal elements). Θ is the alternative-wise gaze discount vector (where the ith entry is set to 1 when item i is fixated, and other entries are set to the discount value θ, $0 \leq \theta \leq 1$). C is a contrast matrix which, as in MDFT, can perform comparisons between the entries of the item value vector x. We now describe the different nodes and levels of the analysis, that are combined to generate the different model variants:

Attribute integration

The attribute integration switch had two levels: First, outcome probability and outcome value could be integrated *multiplicatively*, so that expected outcome values per item are constructed. This level included the probability weighting function w (Eq. (10)) using a free parameter γ ($0.28 \leq \gamma \leq 1$), and a utility function U (Eq. (3) free parameter α ($0 \leq \alpha$). The item values are given as

$$x_i = w(p_i)U(m_i). \tag{14}$$

Alternatively, attribute integration could be implemented in a *weighted additive* fashion[43]. In this case, attributes were first normalized using divisive normalization[73] to make them commensurable on a single scale:

$$p_i^{norm} = \frac{p_i}{\sum_i^n p_i} \tag{15}$$

and

$$m_i^{norm} = \frac{m_i}{\sum_i^n m_i}. \tag{16}$$

Next, the normalized attributes would be combined *additively*, with weighting parameter w_p ($0 \leq w_p \leq 1$), controlling the relative contribution of the probability attribute:

$$x_i = w_p p_i^{norm} + (1 - w_p) m_i^{norm} \tag{17}$$

Evidence comparison

The evidence comparison switch had two levels: First, item values x_i could be accumulated *independently* for each alternative, without comparison to other alternatives. In this case the contrast matrix C is set to an identity matrix. Second, item values x_i could be accumulated in a *comparative* fashion. Then, the contrast matrix C is set up to perform comparisons between each input x_i and the mean of all other inputs $x_{j \neq i}$, as in MDFT. To this end, diagonal entries of C are set to 1, and off-diagonal elements to $\frac{-1}{N-1}$, where N is the number of alternatives (here $N = 3$).

Alternative-wise gaze discount

This switch could take the values *"on"* or *"off"*. If switched on, the model included a free parameter θ ($0 \leq \theta \leq 1$) controlling the discount rate of unattended alternatives during accumulation. If switched off, the gaze discount vector Θ was set to one.

Attribute-wise gaze discount

The analysis also included the option of attribute-wise gaze-dependent discounting (similar to the two-layer model from Glickman *et al.*[1] and the model presented in Fisher[41]. If switched on, stimulus attributes of the currently unattended dimension (e.g., probability, when an lottery outcome was fixated) were discounted by a free parameter η ($0 \leq \eta \leq 1$). In combination with *additive* attribute integration, the attribute-wise gaze discount was applied after attribute normalization, but prior to the weighted addition. For *multiplicative* attribute integration, attributes were discounted before entering probability-weighting and utility functions w and U.

Accumulation leak

We investigated three different forms of accumulation leak: First, accumulation without leak. In these variants, the diagonal elements of the feedback matrix S were set to 1, resulting in no leak. The second possibility was uniform *constant* leak, where we estimated a parameter λ ($0 \leq \lambda \leq 1$, where 1 indicates perfect memory without leak, and 0 indicates leak of all prior information), occupying the diagonal elements of the feedback matrix S. The third type of leak we investigated was *gaze-dependent*. Here, only accumulators of unattended alternatives leak according to the λ parameter.

Inhibition

We considered four types of inhibition between accumulators: First, independent accumulation without inhibition. In this case, all off-diagonal elements of S were set to 0. Second, we considered uniform *constant* inhibition, where we estimated a parameter ϕ ($0 \leq \phi \leq 1$) and set each off-diagonal element in S to $-\phi$, resulting in uniform inhibition (proportional to the accumulators' activation level), across items. Thirdly, we considered *distance dependent* inhibition, as implemented in MDFT (see Eq. (7)). Here, the inhibition between accumulators is a function of the corresponding items' distance in attribute space. The distance is expressed in indifference and dominance directions, and the dominance direction is overweighted by a parameter w_d. As we did for MDFT, we log-transformed probability and outcome attributes and rescaled them to a range between 0 and 1 before applying the distance function. This implementation uses a sensitivity parameter ϕ, a parameter estimating the relative importance of the probability attribute w_p (this parameter is already used if attribute integration is *additive*), and the overweighting parameter of the dominance direction w_d. Note, we only computed off-diagonal elements of S according to Eq. (7), as the diagonal entries were controlled by the accumulation leak switch. Finally, we considered *gaze-dependent* inhibition. Here, the rationale is that only the accumulator of the currently attended item exerts inhibition over all others. In this type of inhibition, all off-diagonal elements of S in the column corresponding to the currently attended item are set to $-\phi$, and others are set to 0.

Total number of variants and parameter estimation

Exhaustive combination of all switch levels yields 192 model variants. The effective number of uniquely identifiable models was, however, reduced to 160 because for some variants comparative and independent accumulation versions cannot be distinguished when choice probabilities are derived from a soft-max choice rule with a freely estimated inverse temperature parameter over final accumulator values. This is the case for variants with no or constant inhibition and leak. Each variant was fit individually to the data from each participant by maximum-likelihood estimation, using a Differential Evolution optimization algorithm (see above). As the number of parameters differ between model variants, we computed the BIC for each model and participant to obtain a measure of model fit, corrected for model complexity.

The optimization algorithm failed to find a solution better than a random model for 108 of 6400 (1.69%) of estimation runs. Since all model variants used the soft-max choice rule (Eq. (4)) and therefore could predict random choices by setting the inverse temperature parameter β to 0, this indicates non-convergence of the optimization algorithm. All but one non-converged estimation run used distance-dependent inhibition. We set the maximum-likelihood estimates of the failed runs to that of the nested random model for all analyses.

Statistical modelling

We used Bayesian estimation (BEST)[32,33] of differences, effect size d and associated 95% highest posterior density intervals (HDI_{95}) for all reported pairwise comparisons. Reported correlation coefficients and associated HDI_{95} result from Bayesian correlation analyses[74]. Regression estimates and HDI_{95} result from Bayesian regression analyses implemented in PyMC3[75].

Software

The task was programmed in MATLAB (The Mathworks Inc., USA) using the PsychToolBox[76]. Data processing and analyses were done in Python with numpy[77], scipy[72] and pandas[78] libraries. Bayesian analyses were implemented in PyMC3[75], mixed models used bambi[79]. Exceedance probabilities were computed in MATLAB using SPM12[80]. Fig.s were created using matplotlib[81], seaborn[82] and python-ternary[83].

Data and code availability

All raw and preprocessed data, and scripts to reproduce all reported processing, analyses and figures are available at https://github.com/moltaire/gda-context.

Acknowledgements

F.M. is supported by the International Max Planck Research School on the Life Course (LIFE). A.T. is supported by the Max Planck School of Cognition and Stanford Data Science.

Author contributions

F.M., S.A.H. and P.N.C.M. designed the study. F.M. and A.W.T. conceptualized the behavioural models. F.M. performed the experiment, wrote all software, analysed the data and visualized results. P.N.C.M. and H.R.H. supervised the project. P.N.C.M. acquired funding. F.M. wrote the original draft. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Glickman, M. *et al.* The Formation of Preference in Risky Choice. *PLOS Computational Biology* **15,** e1007201 (2019).
2. Gluth, S., Spektor, M. S. & Rieskamp, J. Value-Based Attentional Capture Affects Multi-Alternative Decision Making. *eLife* **7** (ed Frank, M. J.) e39659 (2018).
3. Smith, S. M. & Krajbich, I. Attention and Choice across Domains. *Journal of Experimental Psychology: General* **147,** 1810–1826 (2018).
4. Armel, K. C., Beaumel, A. & Rangel, A. Biasing Simple Choices by Manipulating Relative Visual Attention. *Judgment and Decision Making* **3,** 8 (2008).
5. Cavanagh, J. F., Wiecki, T. V., Kochar, A. & Frank, M. J. Eye Tracking and Pupillometry Are Indicators of Dissociable Latent Decision Processes. *Journal of Experimental Psychology. General* **143,** 1476–1488 (2014).
6. Krajbich, I., Armel, C. & Rangel, A. Visual Fixations and the Computation and Comparison of Value in Simple Choice. *Nature Neuroscience* **13,** 1292–1298 (2010).
7. Pärnamets, P. *et al.* Biasing Moral Decisions by Exploiting the Dynamics of Eye Gaze. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 4170–4175 (2015).
8. Shimojo, S., Simion, C., Shimojo, E. & Scheier, C. Gaze Bias Both Reflects and Influences Preference. *Nature Neuroscience* **6,** 1317–1322 (2003).
9. Sui, X.-Y., Liu, H.-Z. & Rao, L.-L. The Timing of Gaze-Contingent Decision Prompts Influences Risky Choice. *Cognition* **195,** 104077 (2020).
10. Tavares, G., Perona, P. & Rangel, A. The Attentional Drift Diffusion Model of Simple Perceptual Decision-Making. *Frontiers in Neuroscience* **11,** 468 (2017).

11. Krajbich, I. & Rangel, A. Multialternative Drift-Diffusion Model Predicts the Relationship between Visual Fixations and Choice in Value-Based Decisions. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 13852–13857 (2011).

12. Molter, F., Thomas, A. W., Heekeren, H. R. & Mohr, P. N. C. GLAMbox: A Python Toolbox for Investigating the Association between Gaze Allocation and Decision Behaviour. *PLoS ONE* **14** (2019).

13. Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R. & Mohr, P. N. C. Gaze Bias Differences Capture Individual Choice Behaviour. *Nature Human Behaviour* **3,** 625–635 (2019).

14. Von Neumann, J. & Morgenstern, O. *Theory of Games and Economic Behavior* 2. ed. (Princeton University Press, Princeton, 1947).

15. Kahneman, D. & Tversky, A. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47,** 263–291 (1979).

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

17. Luce, R. D. *Individual Choice Behavior* (John Wiley, Oxford, England, 1959).

18. Huber, J., Payne, J. W. & Puto, C. Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research* **9,** 90–98 (1982).

19. Simonson, I. Choice Based on Reasons: The Case of Attraction and Compromise Effects. *Journal of Consumer Research* **16,** 158–174 (1989).

20. Tversky, A. Elimination by Aspects: A Theory of Choice. *Psychological Review* **79,** 281–299 (1972).

21. Mohr, P. N. C., Heekeren, H. R. & Rieskamp, J. Attraction Effect in Risky Choice Can Be Explained by Subjective Distance Between Choice Alternatives. *Scientific Reports* **7,** 1–10 (2017).

22. Soltani, A., De Martino, B. & Camerer, C. A Range-Normalization Model of Context-Dependent Choice: A New Model and Evidence. *PLoS Computational Biology* **8** (2012).

23. Wedell, D. H. Distinguishing among Models of Contextually Induced Preference Reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **17,** 767–778 (1991).

24. Rieskamp, J., Busemeyer, J. R. & Mellers, B. A. Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice. *Journal of Economic Literature* **44,** 631–661 (2006).

25. Roe, R. M., Busemeyer, J. R. & Townsend, J. T. Multialternative Decision Field Theory: A Dynamic Connectionst Model of Decision Making. *Psychological Review* **108,** 370–392 (2001).

26. Trueblood, J. S., Brown, S. D. & Heathcote, A. The Multiattribute Linear Ballistic Accumulator Model of Context Effects in Multialternative Choice. *Psychological Review* **121,** 179–205 (2014).

27. Usher, M. & McClelland, J. L. Loss Aversion and Inhibition in Dynamical Models of Multialternative Choice. *Psychological Review* **111,** 757–769 (2004).

28. Tversky, A. & Simonson, I. Context-Dependent Preferences. *Management Science* **39,** 1179–1189 (1993).

29. Marini, M., Ansani, A. & Paglieri, F. Attraction Comes from Many Sources: Attentional and Comparative Processes in Decoy Effects. *Judgment and Decision Making* **15,** 23 (2020).

30. Cohen, A. L., Kang, N. & Leise, T. L. Multi-Attribute, Multi-Alternative Models of Choice: Choice, Reaction Time, and Process Tracing. *Cognitive Psychology* **98,** 45–72 (2017).

31. Berkowitsch, N. A. J., Scheibehenne, B. & Rieskamp, J. Rigorously Testing Multialternative Decision Field Theory against Random Utility Models. *Journal of Experimental Psychology: General* **143,** 1331–1348 (2014).

32. Kruschke, J. K. Bayesian Estimation Supersedes the t Test. *Journal of Experimental Psychology: General* **142,** 573–603 (2013).

33. Kruschke, J. K. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan* ISBN: 978-0-12-405916-0 (Academic Press, 2014).

34. Trueblood, J. S. Multialternative Context Effects Obtained Using an Inference Task. *Psychonomic Bulletin & Review* **19,** 962–968 (2012).

35. Trueblood, J. S., Brown, S. D. & Heathcote, A. The Fragile Nature of Contextual Preference Reversals: Reply to Tsetsos, Chater, and Usher (2015). *Psychological Review* **122,** 848–853 (2015).

36. Trueblood, J. S., Brown, S. D., Heathcote, A. & Busemeyer, J. R. Not Just for Consumers: Context Effects Are Fundamental to Decision Making. *Psychological Science* **24,** 901–908 (2013).

37. Schwarz, G. Estimating the Dimension of a Model. *The Annals of Statistics* **6,** 461–464 (1978).

38. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian Model Selection for Group Studies — Revisited. *NeuroImage* **84,** 971–985 (2014).

39. Turner, B. M., Schley, D. R., Muller, C. & Tsetsos, K. Competing Theories of Multialternative, Multiattribute Preferential Choice. *Psychological Review* **125,** 329–362 (2018).

40. Ashby, N. J. S., Jekel, M., Dickert, S. & Glöckner, A. Finding the Right Fit: A Comparison of Process Assumptions Underlying Popular Drift-Diffusion Models. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **42,** 1982–1993 (2016).

41. Fisher, G. An Attentional Drift Diffusion Model over Binary-Attribute Choice. *Cognition* **168,** 34–45 (2017).

42. Krajbich, I., Lu, D., Camerer, C. & Rangel, A. The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions. *Frontiers in Psychology* **3,** 193 (2012).

43. Rouault, M., Drugowitsch, J. & Koechlin, E. Prefrontal Mechanisms Combining Rewards and Beliefs in Human Decision-Making. *Nature Communications* **10,** 1–16 (2019).

44. Busemeyer, J. R., Gluth, S., Rieskamp, J. & Turner, B. M. Cognitive and Neural Bases of Multi-Attribute, Multi-Alternative, Value-Based Decisions. *Trends in Cognitive Sciences* **23,** 251–263 (2019).

45. Liew, S. X., Howe, P. D. L. & Little, D. R. The Appropriacy of Averaging in the Study of Context Effects. *Psychonomic Bulletin & Review* **23,** 1639–1646 (2016).

46. Fiedler, S. & Glöckner, A. The Dynamics of Decision Making in Risky Choice: An Eye-Tracking Analysis. *Frontiers in Psychology* **3** (2012).

47. Glöckner, A. & Herbold, A.-K. An Eye-Tracking Study on Information Processing in Risky Decisions: Evidence for Compensatory Strategies Based on Automatic Processes. *Journal of Behavioral Decision Making* **24,** 71–98 (2011).

48. Johnson, E. J., Schulte-Mecklenbeck, M. & Willemsen, M. C. Process Models Deserve Process Data: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review* **115,** 263–272 (2008).

49. Russo, J. E. & Dosher, B. A. Strategies for Multiattribute Binary Choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **9,** 676–696 (1983).

50. Su, Y. *et al.* Is Making a Risky Choice Based on a Weighting and Adding Process? An Eye-Tracking Investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **39,** 1765–1780 (2013).

51. Amasino, D. R., Sullivan, N. J., Kranton, R. E. & Huettel, S. A. Amount and Time Exert Independent Influences on Intertemporal Choice. *Nature Human Behaviour* **3,** 383–392 (2019).

52. Noguchi, T. & Stewart, N. In the Attraction, Compromise, and Similarity Effects, Alternatives Are Repeatedly Compared in Pairs on Single Dimensions. *Cognition* **132,** 44–56 (2014).

53. Cataldo, A. M. & Cohen, A. L. The Comparison Process as an Account of Variation in the Attraction, Compromise, and Similarity Effects. *Psychonomic Bulletin & Review* **26,** 934–942 (2019).

54. Folke, T., Jacobsen, C., Fleming, S. M. & Martino, B. D. Explicit Representation of Confidence Informs Future Value-Based Decisions. *Nature Human Behaviour* **1,** 1–8 (2016).

55. Gluth, S., Kern, N., Kortmann, M. & Vitali, C. L. Value-Based Attention but Not Divisive Normalization Influences Decisions with Multiple Alternatives. *Nature Human Behaviour,* 1–12 (2020).

56. Towal, R. B., Mormann, M. & Koch, C. Simultaneous Modeling of Visual Saliency and Value Computation Improves Predictions of Economic Choice. *Proceedings of the National Academy of Sciences* **110,** E3858–E3867 (2013).

57. Vaidya, A. R. & Fellows, L. K. Testing Necessary Regional Frontal Contributions to Value Assessment and Fixation-Based Updating. *Nature Communications* **6,** 10120 (2015).

58. Stewart, N., Hermens, F. & Matthews, W. J. Eye Movements in Risky Choice: Eye Movements in Risky Choice. *Journal of Behavioral Decision Making* **29,** 116–136 (2016).

59. Liu, H.-Z. *et al.* Exploiting the Dynamics of Eye Gaze to Bias Intertemporal Choice. *Journal of Behavioral Decision Making* **n/a** (2020).

60. Evans, N. J., Holmes, W. R., Dasari, A. & Trueblood, J. S. The Impact of Presentation Order on Attraction and Repulsion Effects in Decision-Making. *Decision* **8,** 36–54 (2021).

61. Bhatia, S. Associations and the Accumulation of Preference. *Psychological Review* **120,** 522–543 (2013).

62. Noguchi, T. & Stewart, N. Multialternative Decision by Sampling: A Model of Decision Making Constrained by Process Data. *Psychological Review* **125,** 512–544 (2018).

63. Wollschläger, L. M. & Diederich, A. The 2N-Ary Choice Tree Model for N-Alternative Preferential Choice. *Frontiers in Psychology* **3** (2012).

64. Tsetsos, K., Chater, N. & Usher, M. Salience Driven Value Integration Explains Decision Biases and Preference Reversal. *Proceedings of the National Academy of Sciences* **109,** 9659–9664 (2012).

65. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* Second edition. ISBN: 978-0-262-03924-6 (The MIT Press, Cambridge, Massachusetts, 2018).

66. Hotaling, J. M., Busemeyer, J. R. & Li, J. Theoretical Developments in Decision Field Theory: Comment on Tsetsos, Usher, and Chater (2010). *Psychological Review* **117,** 1294–1298 (2010).

67. Busemeyer, J. R. & Diederich, A. Survey of Decision Field Theory. *Mathematical Social Sciences. Random Utility Theory and Probabilistic Measurement Theory* **43,** 345–370 (2002).

68. Böckenholt, U. in *Multidimensional Models of Perception and Cognition* 89–113 (Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1992). ISBN: 978-0-8058-0577-2.

69. Usher, M. & McClelland, J. L. The Time Course of Perceptual Choice: The Leaky, Competing Accumulator Model. *Psychological Review* **108,** 550–592 (2001).

70. Cavagnaro, D. R., Pitt, M. A., Gonzalez, R. & Myung, J. I. Discriminating among Probability Weighting Functions Using Adaptive Design Optimization. *Journal of Risk and Uncertainty* **47,** 255–289 (2013).

71. Storn, R. & Price, K. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization* **11,** 341–359 (1997).

72. Jones, E., Oliphant, T. & Peterson, P. SciPy: Open Source Scientific Tools for Python (2001).

73. Louie, K., Khaw, M. W. & Glimcher, P. W. Normalization Is a General Neural Mechanism for Context-Dependent Decision Making. *Proceedings of the National Academy of Sciences* **110,** 6139–6144 (2013).

74. Lee, M. D. & Wagenmakers, E.-J. *Bayesian Cognitive Modeling: A Practical Course* ISBN: 978-1-107-01845-7 978-1-107-60357-8 (Cambridge University Press, Cambridge ; New York, 2013).

75. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic Programming in Python Using PyMC3. *PeerJ Computer Science* **2,** e55 (2016).

76. Brainard, D. H. The Psychophysics Toolbox. *Spatial vision* **10,** 433–436 (1997).

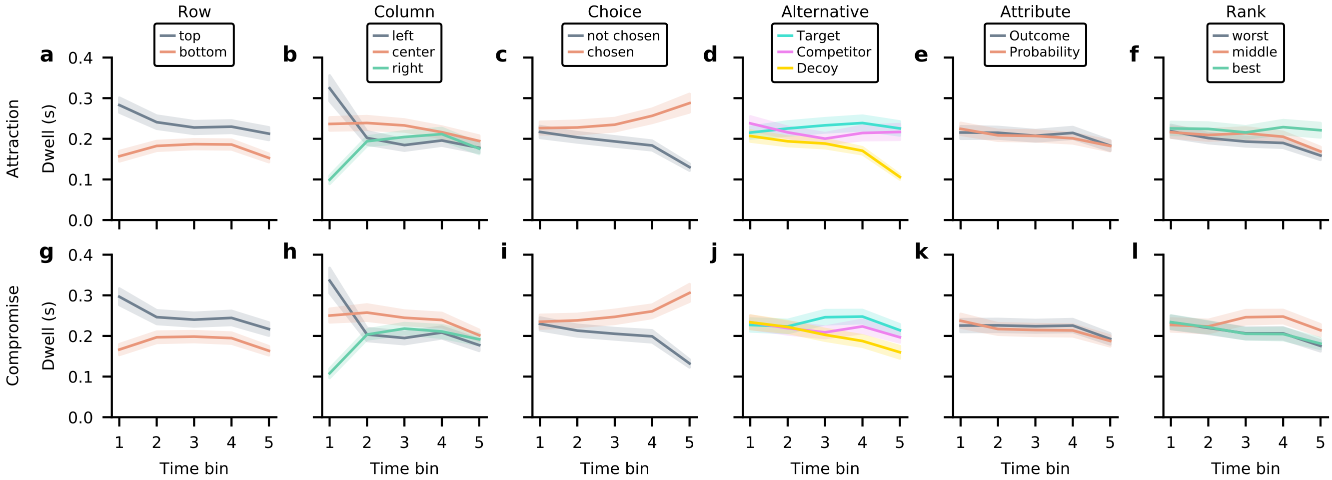
77. Harris, C. R. *et al.* Array Programming with NumPy. *Nature* **585,** 357–362 (2020).

78. McKinney, W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* ISBN: 978-1-4493-2361-5 ("O'Reilly Media, Inc.", 2012).

79. Yarkoni, T. & Westfall, J. Bambi: A Simple Interface for Fitting Bayesian Mixed Effects Models (2016).
80. Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J. & Nichols, T. E. *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (Elsevier, 2011).
81. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9,** 90–95 (2007).
82. Waskom, M. & seaborn development team, t. *Mwaskom/Seaborn* Zenodo. 2020.
83. Harper, M. *et al.* Python-Ternary: Ternary Plots in Python. *Zenodo 10.5281/zenodo.594435* (2015).

Supplementary Information

Supplementary Note 1

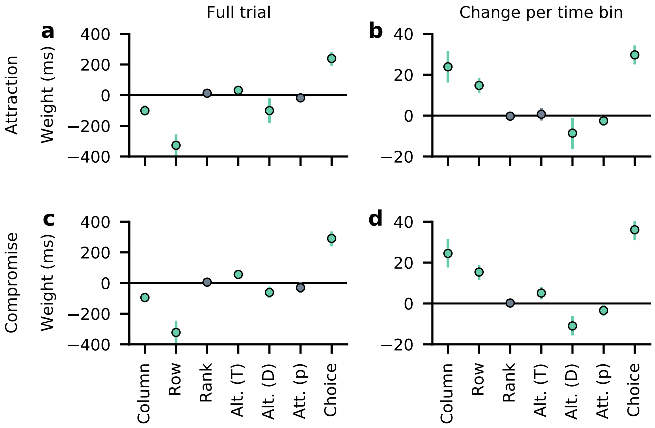


Supplementary Fig. 1. Distribution of gaze over the course of the trial depending on stimulus characteristics. Each panel shows the average dwell time towards AOIs for a given stimulus feature (e.g., horizontal and vertical position) across five time-bins. Data is shown separately for attraction **(a-f)** and compromise **(g-l)** trials.

Regression analyses of gaze behaviour

We performed two linear mixed effects regressions of total dwell time towards an AOI in each trial, separate for attraction and compromise trials (Supplemental Fig. 1). In the first ("full trial") model, the dependent variable was total dwell time towards an AOI across the full trial. This model used the following predictors: Vertical position (row, centred: -0.5 = top row, +0.5, bottom row), horizontal position (column, centred: -1 = left, 0 = centre, +1 = right), attribute (dummy coding probability attribute p), within dimension attribute rank (centred, -1 = worst, 0 = intermediate, +1 = best on attribute), target (dummy coded), decoy (dummy coded) and a dummy coded predictor for the ultimately chosen alternative. For the second model, we partitioned the dwell-data into five equally sized time-bins. The dependent variable in this model then was total dwell time towards an AOI within a time-bin. This model included the same predictors as the "full trial" model. Crucially, it also included interaction terms for each predictor with the time-bin variable, and time-bin as additional predictor. Both models included random intercepts and slopes for each participant. Bayesian posterior distributions of the regression weights were estimated using the bambi Python library[1], with default priors[2], sampling four chains with 2000 samples each, after a tuning phase of 500 samples. Convergence was diagnosed visually and by means of the Gelman-Rubin statistic ($|1 - \hat{R}| \leq 0.05$ for all chains).

Regression weight estimates are shown in Supplemental Fig. 2. Across the trial, we find strong effects of position on dwell time, so that dwell times towards the top and left were longer. Significant negative interaction effects of the row and column predictors with time showed that these effects diminish across the trial. We also find a gaze-cascade effect[3,4], where dwell times to AOIs belonging to the ultimately chosen alternative are longer across the trial (and increasingly so throughout the trial, indicated by the positive interaction term with time). Dwell times to decoys decreased significantly during the trial, and across the full trial, dwell times to decoys were shorter than other alternatives. Similarly, dwell times towards probability attributes p shortened across the trial. Across the full trial, however, dwell times towards probability AOIs were not shorter than those to outcome AOIs. Finally, dwell times to target alternatives were longer across the trial in both compromise and attraction trials. In addition, this effect increased throughout the trial in compromise, but not in attraction trials. Note that these effects are



Supplementary Fig. 2. Weight estimates from regression analyses on absolute dwell times. We performed two mixed-effects regression analyses of dwell time towards each AOI onto stimulus properties: **(a, c)** Regressing the total dwell time towards an AOI over a full trial onto AOI column, row, attribute rank (best, middle or worst value on the attribute), two dummy predictors coding alternative, attribute (probability or outcome) and whether the AOI belonged to the subsequently chosen alternative. **(b, d)** Second, we binned dwell times in each trial into five time-bins and added an interaction term with time-bin for each predictor. The panels show the interaction term weights. Analyses were carried out separately for attraction **(a, b)** and compromise **(c, d)** trials. Regression models had random intercepts and slopes across participants. Points and intervals mark posterior mean estimates and associated HDI_{95} (coloured green if the interval excluded 0).

independent of the effect of choice, as choice is a separate predictor in the model. We could not find an association between the attribute rank (being the worst, best, or intermediate item on an attribute) and dwell time.

Direction of information search

We further analysed participants' direction of information search. Therefore, we counted the number of vertical (transitions within the same alternative), horizontal (within the same row, between alternatives) and diagonal (between rows and alternatives) transitions. On average, participants made over 7 horizontal transitions in attraction (mean $\pm$ s.d. $= 7.28 \pm 2.36$) and compromise (mean $\pm$ s.d. $= 7.29 \pm 2.63$) trials, with no meaningful difference between effects. Participants made, however, more vertical transitions in compromise trials (mean $\pm$ s.d. $= 7.55 \pm 3.11$) than attraction trials (mean $\pm$ s.d. $= 7.17 \pm 3.06$; mean difference $= 0.39$, $\mathrm{HDI}_{95} = [0.11, 0.64]$, $d = 0.49$, $\mathrm{HDI}_{95} = [0.13, 0.82]$). The number of diagonal transitions was lower overall, but higher in attraction (mean $\pm$ s.d. $= 2.88 \pm 1.18$) than compromise trials (mean $\pm$ s.d. $= 2.72 \pm 1.31$; mean difference $= 0.19$, $\mathrm{HDI}_{95} = [0.06, 0.32]$, $d = 0.62$, $\mathrm{HDI}_{95} = [0.14, 1.17]$).

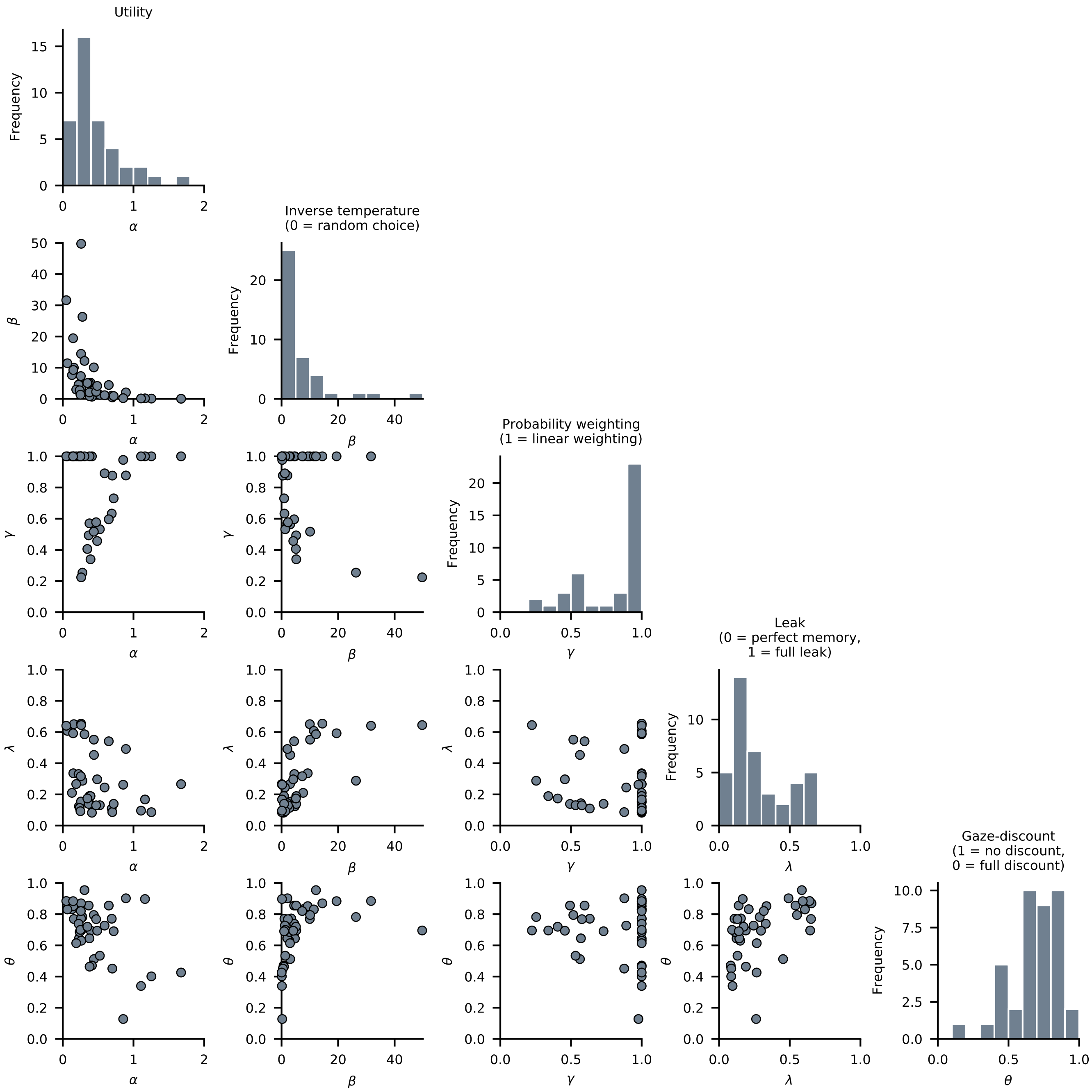
While vertical transitions always translate to transitions "within alternative", horizontal transitions are not necessarily always "within attribute", since the attribute positions of each alternative were random in the task. We therefore recoded transitions as "within alternative", "within attribute" and "between alternatives and attributes" and computed the Payne Index[5] for each trial as:

$$\text{Payne Index} = \frac{N_{\text{within alt.}} - N_{\text{within att.}}}{N_{\text{within alt.}} + N_{\text{within att.}}} \tag{18}$$

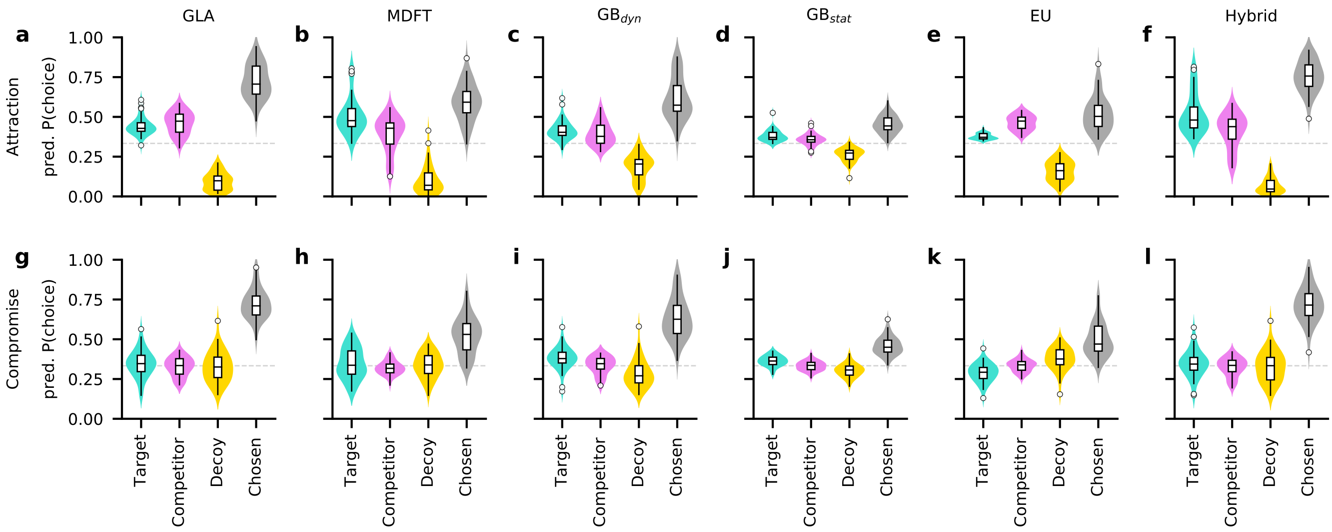
A more positive value on the index indicates more processing within alternatives, whereas more negative values indicate more processing between alternatives, within the same attribute dimension. Overall, the average Payne Index was slightly positive for both attraction (mean $\pm$ s.d. $= 0.10 \pm 0.16$) and compromise trials (mean $\pm$ s.d. $= 0.14 \pm 0.18$), suggesting a mixture of within-alternative and within-attribute processing, with slightly more processing within alternatives. It was, however, lower in attraction trials (mean difference $= 0.04$, $\mathrm{HDI}_{95} = [0.01, 0.06]$, $d = 0.51$, $\mathrm{HDI}_{95} = [0.16, 0.87]$), implying comparably more processing between alternatives in attraction trials.

	mean	SD	min	25%	50%	75%	max
α	0.47	0.35	0.05	0.25	0.37	0.61	1.67
β	6.81	9.87	0.04	1.13	3.25	8.04	49.74
γ	0.81	0.25	0.22	0.58	1.0	1.0	1.0
λ	0.29	0.2	0.08	0.14	0.23	0.46	0.65
θ	0.69	0.18	0.13	0.63	0.72	0.83	0.95

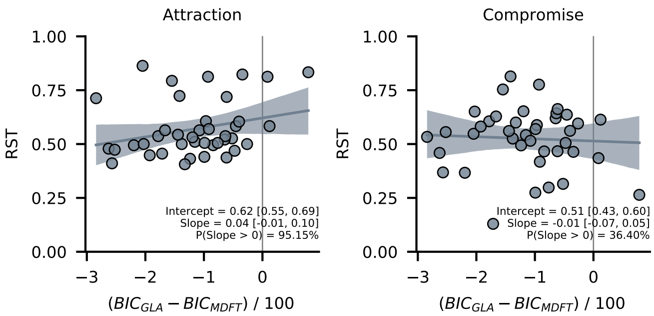
Supplementary Table 1. Summary of GLA estimates. α is the utility parameter. β is the inverse temperature parameter of the choice rule (0 = random choice). γ is the probability weighting parameter (1 = objective probability weighting). λ is the leak parameter (0 = perfect memory, 1 = full leak of all previous information). θ is the gaze-discount parameter (1 = no gaze-discount, 0 = maximum gaze-discount).



Supplementary Fig. 3. GLA maximum likelihood estimates and relationships between parameters. α is the utility parameter. β is the inverse temperature parameter of the choice rule (0 = random choice). γ is the probability weighting parameter (1 = linear weighting). λ is the leak parameter (0 = perfect memory, 1 = full leak of all previous information). θ is the gaze-discount parameter (1 = no gaze-discount, 0 = maximum gaze-discount).



Supplementary Fig. 4. Model-predicted choice probabilities. Each panel shows distributions of participant-level mean model-predicted choice probabilities for the target, competitor, decoy and ultimately chosen alternative. Predictions for attraction and compromise trials are displayed separately in the top **(a-f)** and bottom rows **(g-l)**. Predictions were computed using individual maximum likelihood estimates. The hybrid model **(f, l)** was derived from the switchboard analysis and combines an alternative-wise gaze-discount with a distance-dependent inhibition mechanism. Violin plots show kernel density estimates of distributions of individual values. Box plots mark lower and upper quartiles and median. Whiskers extend from first and last datum within 1.5 times the interquartile range from lower and upper quartiles, respectively. Values outside this range are indicated by open circles.



Supplementary Fig. 5. Relative model fits between GLA and MDFT in relation to RST. Relative model fits of MDFT (indicated by BIC difference between GLA and MDFT) tended to be higher for participants with higher RST in attraction trials (left panel; slope = 0.04, HDI_{95} = [-0.01, 0.09] increase in RST per 100 unit increase in BIC difference, 93.6% of posterior mass above 0), but not compromise trials (right panel), even though 7 out of 9 participants with attraction RST above 0.7 were better described by GLA overall (participants left of dashed vertical line).

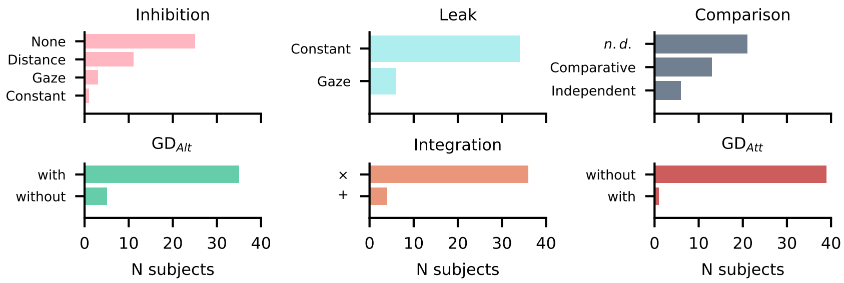
Switchboard analysis overview

Switch	Gaze-*in*dependent levels			Gaze-dependent levels	N
Attribute integration	**× Multiplicative** Outcomes and probabilities combine multiplicatively into expected utilities[6,7].	**+ Weighted additive** Outcomes and probabilities are normalized, weighted and added[8].		*n.a.*	2
Comparison	**Independent** Accumulation of absolute item values per alternative[7,9].	**Comparative** Accumulation of relative values per alternative[10].		*n.a.*	2
Alternative-wise gaze-discount	**False** Fixated and non-fixated alternatives processed equally.			**True** Non-fixated alternatives' values are discounted by a parameter θ[11].	2
Attribute-wise gaze-discount	**False** Fixated and non-fixated attributes are processed equally.			**True** Attributes on the non-fixated dimension are discounted by a parameter η[12,13].	2
Accumulation leak	**None** Perfect integration over fixations.	**Constant** With each fixation, accumulators leak information proportional to their current value, controlled by parameter λ[10,14].		**Gaze-dependent** With each fixation, accumulators of non-fixated alternatives leak information, as in Constant leak[15].	3
Accumulator inhibition	**None** No inhibition between accumulators.	**Constant** Accumulators inhibit each other proportional to their current value, controlled by parameter Φ[16].	**Distance-dependent** Inhibition between accumulators depends on pairwise distance between alternatives in attribute space. Parameters w_d and Φ[10].	**Gaze-dependent** With each fixation, accumulators of non-fixated alternatives are inhibited, proportional to the currently fixated accumulator's value. Controlled by parameter Φ.	4
Total variants					192

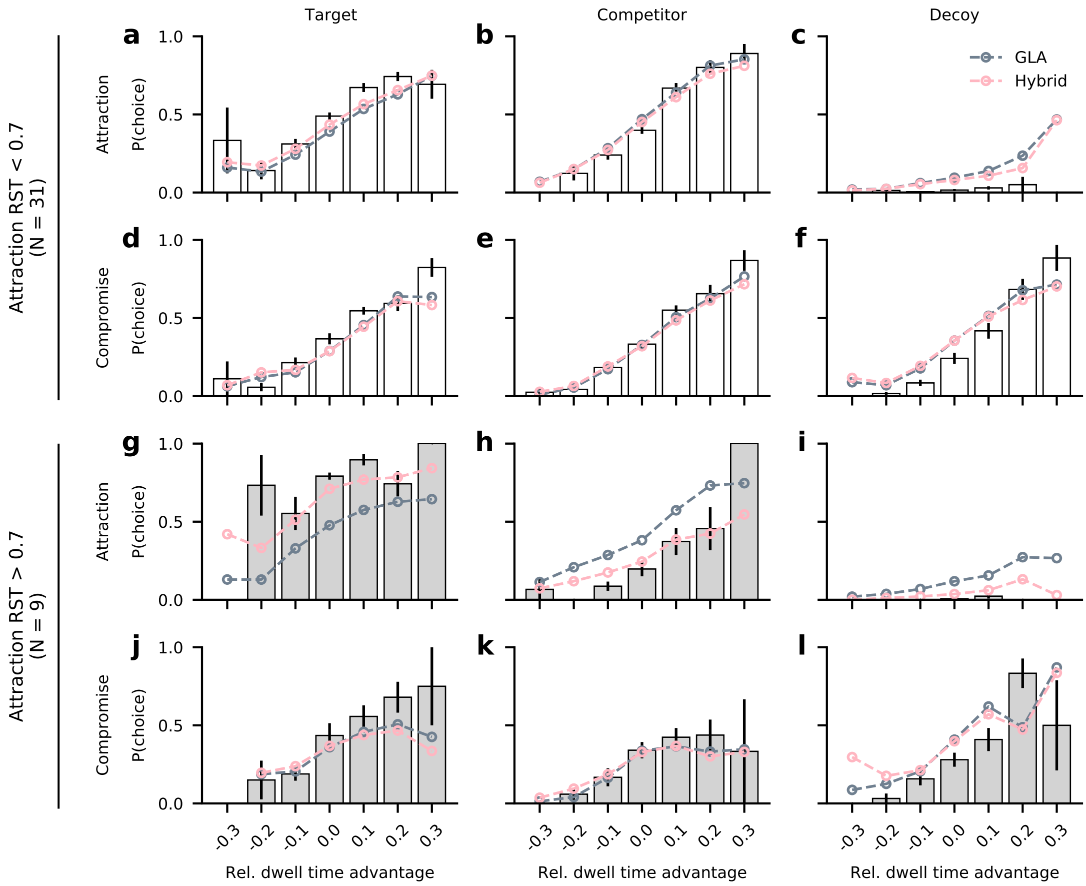
Supplementary Table 2. Overview of the nodes and switch-levels used in the switchboard analysis. Switch-levels that depend on gaze-data are shaded blue. Note that due to the model fitting procedure, where model predicted choice probabilities are derived from a soft-max function over the final accumulator values, the comparison switch levels *independent* and *comparative* are not distinguishable for a subset of the model space, reducing the total number of unique variants to 160.

Rank	$\mathbf{GD_{Alt}}$	$\mathbf{GD_{Att}}$	Leak	Inhibition	Integration	Comparison	BIC
1	**Yes**	No	Constant	None	Multiplicative	*n.d.*	232.08
2	No	No	Constant	**Gaze**	Multiplicative	Independent	235.94
3	**Yes**	No	Constant	**Gaze**	Multiplicative	Comparative	236.75
4	**Yes**	**Yes**	Constant	None	Multiplicative	*n.d.*	237.21
5	**Yes**	No	Constant	**Gaze**	Multiplicative	Independent	237.31
6	**Yes**	No	Constant	Constant	Multiplicative	*n.d.*	237.40
7	**Yes**	No	**Gaze**	None	Multiplicative	Comparative	238.41
8	No	**Yes**	Constant	**Gaze**	Multiplicative	Independent	241.04
9	**Yes**	No	Constant	Distance	Multiplicative	Comparative	241.40
10	**Yes**	**Yes**	Constant	**Gaze**	Multiplicative	Comparative	241.78

Supplementary Table 3. Overview of average best fitting model variants. All ten model variants that fit the data best on average used some form of gaze-dependence (blue shaded cells), mostly an alternative-wise gaze discount. *"n.d."* denotes variants where comparison mechanisms were not distinguishable by the analysis.



Supplementary Fig. 6. Counts of individual best fitting switches. Most participants were best described by model variants that included multiplicative attribute integration, with alternative-wise gaze discount, no attribute-wise gaze discount, constant leakage and no inhibition.



Supplementary Fig. 7. Observed and model-predicted association of dwell time advantage and choice for participants with weaker and strong attraction effects. (a-f) Data and model predictions for participants with weaker attraction effects (RST < 0.7). **(g-l)** Data and model predictions for participants with strong attraction effects (RST > 0.7) Each column refers to one choice alternative: Target (first column; **a, d, g, j**); Competitor (second column; **b, e, h, k**); Decoy (third column; **c, f, i, l**). Rows refer to trials in attraction **(a-c, g-i)** and compromise trials **(d-f, j-l)**. White and grey bars and error bars show observed mean $\pm$ s.e. choice probabilities computed from even-numbered trials, for participants with weaker and stronger attraction effects, respectively. Coloured lines indicate model predictions derived from 50 simulations for each odd-numbered trial.

Supplementary Note 2

No process evidence that strong attraction responders follow simple choice rule

Using process measures, we performed multiple tests of the hypothesis, that individuals with strong attraction effects follow a simple choice rule of choosing the dominant alternative. First, we tested whether the strength of individual attraction effects (individual RST in attraction trials) was related to differences in mean response times (RTs) in attraction trials. If individuals used a choice rule, their choices might be made faster, as they do not engage in multiple pairwise comparisons or calculations of expected outcomes. There was no correlation between the two measures ($r = 0.06$, $\text{HDI}_{95} = [-0.24, 0.34]$). Similarly, no relationship was found between individual RST and the number of fixations in attraction trials ($r = 0.06$, $\text{HDI}_{95} = [-0.25, 0.35]$). Mean RTs in attraction trials did not meaningfully differ between trials with target choices and trials with other choices ($d = -0.2$, $\text{HDI}_{95} = [-0.67, 0.25]$). Next, we tested whether individuals with strong attraction effects committed to a choice once they learned about the dominance relationship in the stimuli, as if using the dominance relationship as a stopping rule, or if they kept exploring the stimuli. There was, however, no relationship between individual RST and the mean number of fixations after all target and decoy attributes were fixated at least once ($r = 0.04$, $\text{HDI}_{95} = [-0.28, 0.31]$). The same analysis using fixation counts after target and decoy alternatives were both seen at least once on any attribute revealed no effect either ($r = 0.11$, $\text{HDI}_{95} = [-0.20, 0.40]$). Taken together, we did not find any evidence based on process data to support the hypothesis that strong attraction responders used a simple choice rule.

Supplementary Information References

1. Yarkoni, T. & Westfall, J. Bambi: A Simple Interface for Fitting Bayesian Mixed Effects Models (2016).
2. Westfall, J. Statistical Details of the Default Priors in the Bambi Library. *arXiv:1702.01201 [stat]* (2017).
3. Mullett, T. L. & Stewart, N. Implications of Visual Attention Phenomena for Models of Preferential Choice. *Decision* **3,** 231 (2016).
4. Shimojo, S., Simion, C., Shimojo, E. & Scheier, C. Gaze Bias Both Reflects and Influences Preference. *Nature Neuroscience* **6,** 1317–1322 (2003).
5. Payne, J. W. Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis. *Organizational Behavior and Human Performance* **16,** 366–387 (1976).
6. Tversky, A. & Kahneman, D. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* **5,** 297–323 (1992).
7. Glickman, M. *et al.* The Formation of Preference in Risky Choice. *PLOS Computational Biology* **15,** e1007201 (2019).
8. Rouault, M., Drugowitsch, J. & Koechlin, E. Prefrontal Mechanisms Combining Rewards and Beliefs in Human Decision-Making. *Nature Communications* **10,** 1–16 (2019).
9. Bhatia, S. Associations and the Accumulation of Preference. *Psychological Review* **120,** 522–543 (2013).
10. Roe, R. M., Busemeyer, J. R. & Townsend, J. T. Multialternative Decision Field Theory: A Dynamic Connectionst Model of Decision Making. *Psychological Review* **108,** 370–392 (2001).
11. Krajbich, I., Armel, C. & Rangel, A. Visual Fixations and the Computation and Comparison of Value in Simple Choice. *Nature Neuroscience* **13,** 1292–1298 (2010).
12. Krajbich, I., Lu, D., Camerer, C. & Rangel, A. The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions. *Frontiers in Psychology* **3,** 193 (2012).
13. Fisher, G. An Attentional Drift Diffusion Model over Binary-Attribute Choice. *Cognition* **168,** 34–45 (2017).

14. Usher, M. & McClelland, J. L. The Time Course of Perceptual Choice: The Leaky, Competing Accumulator Model. *Psychological Review* **108,** 550–592 (2001).

15. Ashby, N. J. S., Jekel, M., Dickert, S. & Glöckner, A. Finding the Right Fit: A Comparison of Process Assumptions Underlying Popular Drift-Diffusion Models. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **42,** 1982–1993 (2016).

16. Usher, M. & McClelland, J. L. Loss Aversion and Inhibition in Dynamical Models of Multialternative Choice. *Psychological Review* **111,** 757–769 (2004).

Molter, F. & Mohr, P. N. C. (submitted). Presentation order but not duration affects binary risky choice. *PsyArXiv*. https://doi.org/10.31234/osf.io/gcthj

Presentation order but not duration affects binary risky choice

Felix Molter[1,2,*] and Peter N. C. Mohr[1,2]

[1]School of Business & Economics, Freie Universität Berlin, Berlin, Germany
[2]Center for Cognitive Neuroscience, Freie Universität Berlin, Berlin, Germany
[*]E-mail: felixmolter@gmail.com

September 1, 2021

Abstract

Risky choice behaviour often deviates from the predictions of normative models. The information search process has been suggested as a source of some reported "biases". Specifically, gaze-dependent evidence accumulation models, where unfixated alternatives' signals are discounted, propose a mechanistic account of observed associations between eye movements, choices and response times, with longer fixated alternatives being chosen more frequently. It remains debated, however, whether gaze causally influences the choice process, or rather reflects emerging preferences. Furthermore, other aspects the information search process, like the order in which information is inspected, can be confounded with gaze duration, complicating the identification of their causal influences. In our preregistered study 179 participants made repeated incentivized choices between two sequentially presented risky gambles, allowing the experimental control of presentation duration, order, and format (i.e., alternative-wise or attribute-wise). Across presentation formats, we find evidence against an influence of presentation duration on choice. The order in which participants were shown stimulus information, however, causally affected choices, with alternatives shown last being chosen more frequently. Notably, while gaze-dependent accumulation models generally capture effects of gaze duration, causal effects of stimulus order are only predicted by some models, identifying potential for future theory development.

Introduction

A large body of experimental research demonstrates that human risky choices systematically differ from those predicted by optimal, utility-maximising and context-invariant theories of choice like Expected Utility Theory (e.g., Allais, 1953; Hertwig et al., 2004; Kahneman & Tversky, 2012; Mohr et al., 2017; Molter et al., 2021). Many descriptive theories of risky choice ascribe these departures to perceptual or attentional processes (e.g., Busemeyer & Townsend, 1993; Tversky & Kahneman, 1992; Yechiam & Hochman, 2013), with the idea that the decision maker's capacity to process information is limited, and attention serves to select and amplify information about the choice setting (e.g., different alternatives' attributes, outcomes, or their probabilities). This selection and weighting of information is then assumed to induce biases in the decision process and ultimately affect the choice itself.

While the specific construct of attention differs between theories and its usefulness as a general explanatory device has been questioned (Hommel et al., 2019), eye movements are often taken as an indicator of which information a decision maker processes during the decision. The most ubiquitous finding in eye tracking studies of decision-making (including risky choices) is that alternatives that are looked at longer are more likely to be chosen (Ashby et al., 2016; Cavanagh et al., 2014; Fiedler & Glöckner, 2012; Glöckner & Herbold, 2011; Gluth et al., 2020; Gluth et al., 2018; Isham & Geng, 2013; Krajbich et al., 2010; Krajbich & Rangel, 2011; Lopez-Persem et al., 2016; Molter et al., 2021; Sepulveda et al., 2020; Smith & Krajbich, 2018; Stewart, Gächter, et al., 2016; Stewart, Hermens, et al., 2016; Thomas et al., 2021; Thomas et al., 2019). Gaze-dependent accumulation models (Glickman et al., 2019; Krajbich et al., 2010; Krajbich et al., 2012; Krajbich & Rangel, 2011) provide a formal account of this *gaze bias* effect and many other details of the empirical association between gaze and choice. They assume that decisions are made by repeated sampling and accumulation of evidence in favour of each alternative until evidence for one alternative reaches a decision threshold. Crucially, accumulation is assumed to depend on gaze allocation, such that information momentarily outside of the decision maker's gaze is discounted. Notably, gaze-dependent accumulation can explain departures from rational choice (Gluth et al., 2018; Molter et al., 2021), providing theories positing broad attentional effects on choice with concrete process-based evidence.

There remains, however, the question of causality: It is still debated to what extent eye movements causally influence the choice process (and visual attention has a causal role in choice and choice biases), or if they merely reflect an emerging choice (Mormann & Russo, 2021; Westbrook et al., 2020). Note that the accuracy with which gaze-dependent accumulation models predict choices and process measures does not address the issue of causality. Even though these models are frequently interpreted to assume a directed effect of gaze on choice, they only formalize the association between gaze and the choice process, without any directional assumption of causality. For example, it would still be possible for a third variable to influence both choices and gaze. Identifying the direction of causality ultimately requires experimental manipulation.

This issue is of great interest, as a causal effect of gaze on choice would imply that decisions could be influenced by irrelevant, external factors that affect the decision makers' gaze.

Prior work has investigated different aspects in which changes to the information search process affect choice behaviour, experimentally controlling the perceptual saliency of attributes or alternatives (Milosavljevic et al., 2012; Weber & Kirsner, 1997), the duration for which alternatives or attributes were seen (Armel et al., 2008; Lim et al., 2011; Liu, Lyu, et al., 2020; Pärnamets et al., 2015; Shimojo et al., 2003; Sui et al., 2020; Tavares et al., 2017), the temporal order in which alternatives are inspected (Liu, Zhou, et al., 2020), or the direction in which decision makers can gather information about alternatives and their attributes (Mittone & Papi, 2020; Reeck et al., 2017):

Gaze-dependent accumulation models of risky choice predict that a longer gaze towards an alternative is associated with a higher probability of choosing it (unless its value is aversive; see Smith & Krajbich, 2019). If the effect of gaze on choice was causal, this would imply that experimentally inducing longer gaze towards one choice alternative should increase its likelihood of being chosen. A recent study found support for this effect in risky choices using a gaze-contingent decision prompt paradigm, where participants are free to inspect choice alternatives, but are prompted to make their decision when their viewing patterns favour a target attribute or alternative (Sui et al., 2020). Convergent findings have been reported in other decision-making domains (Armel et al., 2008; Liu, Lyu, et al., 2020; Pärnamets et al., 2015; Shimojo et al., 2003; Tavares et al., 2017).

Similarly, as predicted by some gaze-dependent accumulation models (Glickman et al., 2019) longer relative gaze towards individual attributes (e.g., outcomes or probabilities) is associated with changes in choice behaviour (Glickman et al., 2019; Kim et al., 2012), and there is initial evidence that the effect is causal, too (Liu, Lyu, et al., 2020). These results are in line with other studies that, while not controlling gaze duration, found that manipulation of attribute salience affected choices accordingly (Weber & Kirsner, 1997).

Apart from the duration for which alternatives and attributes were evaluated, the temporal order in which information is processed has also been associated with choice. In particular, the last fixation before a choice is often directed towards the chosen alternative. Again, however, the causal direction of this association is debated. The gaze cascade theory (Shimojo et al., 2003) argues that this finding results from reciprocal positive feedback loops between valuation and information search processes, such that preferred alternatives are attended more, which in turn increases preference for them. Order effects are also captured by multiple computational theories of decision making, albeit in different ways: In the attentional Drift Diffusion Model (aDDM) (Krajbich et al., 2010; Krajbich & Rangel, 2011) this phenomenon is explained by the fact that fixated alternatives are more likely to cross the threshold, as evidence for other alternatives is discounted. Other than this, however, the model does not assume a special role of the last fixation, or the serial position of any piece of inspected information. Mullett and Stewart (2016) showed that an increasing association between gaze and choice before a decision is made, naturally emerges in some gaze-dependent accumulation models and does not imply preference-driven information search. Nevertheless, Liu, Zhou, et al. (2020) found that manipulating which of two snack food items was shown last before a decision is prompted causally biased choices in favor of this item. Other models (Ashby et al., 2016; Glickman et al., 2019) assume a leaky integration process that weighs information acquired later in the decision more heavily, and thereby predicts recency effects such that later presented information should affect choices more strongly. Empirical evidence for recency effects in risky choices comes from decisions from experience (Hertwig et al., 2004), where participants learn about alternatives' properties by repeated sampling and related work on value integration in a rapidly presented stream of option outcomes (Tsetsos et al., 2012).

Recent work in multiattribute decisions found that the time point at which information about one attribute dimension is considered in a decision affects choice (Amasino et al., 2019; Maier et al., 2020; Sullivan & Huettel, 2021). In contrast to a simple recency effect, however, these models predict that attributes considered earlier than others exert more influence on the choice process, as this also implies a longer total duration of consideration. These results further highlight the important issue that duration and order effects can be closely related (i.e., last seen alternatives are often also seen longer; and both factors are associated with choice) and need to be distinguished carefully.

Here, we add to this body of research and jointly investigate these multiple potentially causal effects of the information search (presentation duration and order) on choice. We performed a preregistered study using a task that involved incentivized choices between two risky gambles, and was designed to investigate the independent contributions of presentation duration and order to and possible interactions in the choice process. Notably, we tested the duration and order effects on two levels, namely, the level of alternatives and the level of attribute dimensions.

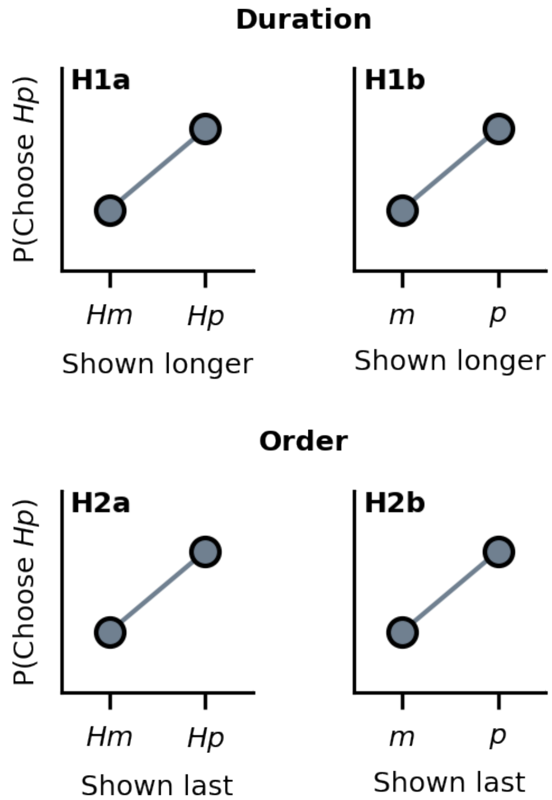


Figure 1. Hypothesized qualitative effects. **H1a**: Higher probability of choosing the alternative with the higher probability (Hp) when it is shown longer. **H1b**: Higher probability of choosing Hp when probabilities are shown longer. **H2a**: Higher probability of choosing Hp when it is shown last. **H2b**: Higher probability of choosing Hp when probabilities are shown last.

Specifically, we hypothesized that presentation *duration* affects choices in two ways: First, in line with gaze-dependent accumulation models, longer presented alternatives should be chosen more frequently (Figure 1 H1a). In addition, we hypothesized that alternatives with higher values on longer presented attributes are chosen more frequently (Figure 1 H1b). Analogously, we hypothesized that presentation *order* affects choices, such that alternatives presented last (Figure 1 H2a), and alternatives with higher values on the attribute presented last (Figure 1 H2b) are chosen more frequently. Our preregistration contained an additional hypothesis and corresponding analysis concerning presentation-dependent changes in value-integration. Details are reported in the Supplementary Information.

In contrast to our predictions and prior literature, our data supports null effects of presentation duration in both attribute- and alternative-wise presentation formats. Instead, we find a causal effect of presentation order on choice, such that alternatives shown last were more likely to be chosen. We discuss the implications of these results.

Results

Behavioural task

In our experiment, 179 participants performed a binary sequential presentation risky choice task (Figure 2), where they made repeated choices between two all-or-nothing risky gambles: One alternative offering a high chance to win a smaller amount (Hp) and a second alternative offering a higher amount with a lower probability (Hm). Each trial consisted of a presentation phase, where participants learned about the two alternatives' winning probabilities p and outcomes m, and a choice phase. Information was presented across four stages either alternative-wise (Figure 2a) or attribute-wise (Figure 2b). We experimentally controlled presentation duration and order in the task: In each trial one alternative (in alternative-wise presentation) or attribute (in attribute-wise presentation) was shown longer than its competitor (3 vs. 2 seconds across the four stages). In addition, alternatives or attributes could be shown first and third, or second and last in the sequence. In total, participants performed 120 experimental trials based on 15 core choice problems (Figure 2c, d), and 20 additional catch trials with a dominant alternative (Figure 2e). See Methods for additional details on the behavioural task and stimuli.

Choice behaviour summary

Participants successfully avoided choosing dominated alternatives in catch trials (average ± s.d. count of dominated choices = 0.35 ± 0.83) with a majority of 140 participants (79%) never choosing a dominated alternative.

In experimental trials, only three participants (1.7%) exclusively chose the Hp (two participants) or Hm (one participant) alternatives. All other participants alternated between alternatives to varying degrees, indicating that the selected choice problems covered individual indifference points between high and low probability gambles for a majority of participants (overall mean ± s.d. P(choose Hp) = 0.62 ± 0.20; Table 1).

Presentation by	Alternatives		Attributes		All
Duration favours	Hm	Hp	Hm	Hp	
Last stage favours					
Hm	0.60	0.60	0.61	0.62	0.61
Hp	0.64	0.63	0.64	0.63	0.63
All	0.62	0.62	0.63	0.62	0.62

Table 1. Mean choice probabilities for the high probability alternative Hp across conditions.

Regression analysis: Effect of presentation order, but not duration

We first analysed choice behaviour using a Bayesian mixed-effects logistic regression model, predicting choice from differences in expected value and effect-coded predictors indicating presentation format, the alternative favoured by presentation duration, the alternative favoured by the last presentation stage, and interaction terms of the duration and order effects with the presentation format (see Methods; Table 2).

Choices were strongly driven by the difference in expected value ($\beta = 1.96$ $[1.77, 2.15]$) such that alternatives with higher expected values were preferred. In addition, the main effect of presentation order was credibly positive ($\beta = 0.17$ $[0.09, 0.25]$) indicating a preference towards alternatives favoured by information shown last. No other main or interaction effect was credibly different from zero (presentation duration $\beta = -0.02$ $[-0.10, 0.05]$; presentation format $\beta = -0.06[-0.13, 0.01]$; duration-by-format interaction $\beta = 0.01$ $[-0.13, 0.16]$; order-by-format interaction $\beta = 0.09$ $[-0.06, 0.23]$).

Next, we addressed our hypotheses regarding effects of presentation duration and order on choice separately, using preregistered directed Bayes factor t-tests and BEST analyses.

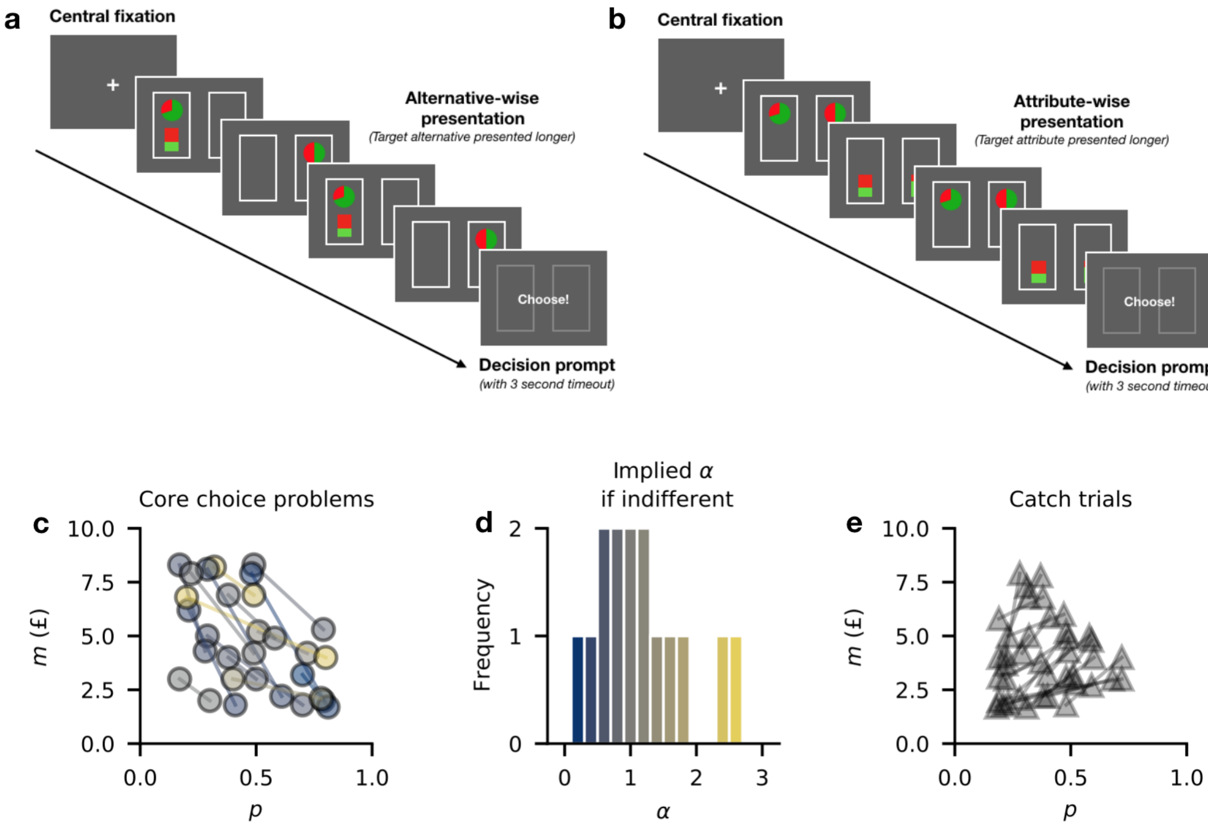


Figure 2. Sequential presentation risky choice task and choice problems. In each trial, participants decided between two all-or-nothing gambles representing the probability p to win an amount m. Information was presented sequentially across four stages so that total presentation duration and presentation order of alternatives (in alternative-wise presentation; **a)** or attributes (in attribute-wise presentation; **b)**) was controlled. After the presentation, participants had 3 seconds to indicate their choice. In each trial, the target alternative or attribute was shown longer than the other. **c)** 15 core choice problems used to construct choice trials. Each circle represents one choice alternative, described by its probability p and outcome m. Each connected pair of circles indicates one pair of choices used to construct the 120 experimental trials. Color indicates the α value for which the expected utilities of a pair are equal (see panel **d**). **d)** Distribution of indifference implied α values of the 15 core choice problems from **c**). **e)** Catch trials. Each connected pair of triangles represents one of 20 catch trials, where one alternative is dominated by the other on both attributes.

No effects of presentation duration across presentation formats

We hypothesized that alternatives shown longer (in alternative-wise presentation) and alternatives with better values on the attribute shown longer (in attribute-wise presentation) would be chosen more frequently. To this end, we computed individuals' probabilities of choosing Hp when it was favoured by presentation duration (because either the alternative Hp or the attribute p was shown longer) and when it was not. We then used directed Bayes Factor t-tests to test the hypotheses of positive versus a null effects in the difference of choice probabilities (see Methods).

On average and across presentation formats, participants chose Hp with a probability of 62.0% when it was favoured by presentation duration, and 62.3% when it was not. There was strong evidence against an effect of presentation duration on choice probability across presentation formats (mean difference = -0.2% $[-1.1\%, 0.6\%]$; mean $d = -0.04$ $[-0.20, 0.11]$; $BF_{+0} = 0.053$, $BF_{0+} = 18.84$). Separate tests for each presentation format confirmed evidence against an effect of presentation duration on choice probability, with strong evidence against an effect in alternative-wise presentation (H1a; mean difference = -0.1% $[-1.5\%, 1.1\%]$; mean $d = -0.02$ $[-0.18, 0.13]$; $BF_{0+} = 15.43$; Figure 3a) and strong evidence against an effect in attribute-wise presentation (H1b; mean

Term	Mean	SD	95% HDI Lower	95% HDI Upper	$\hat{R}$	ESS (bulk)	ESS (tail)
Intercept	0.35	0.11	0.12	0.56	1.01	351	586
EV diff. ($Hp - Hm$; z-scored)	1.96	0.10	1.77	2.15	1.00	763	1962
Duration (Hp or p longer)	−0.02	0.04	−0.10	0.05	1.00	8644	6667
Order (Hp or p last)	0.17	0.04	0.09	0.25	1.00	7602	6943
Format (by alt.)	−0.06	0.04	−0.13	0.01	1.00	10598	6192
Duration × Format	0.01	0.08	−0.13	0.16	1.00	9983	6276
Order × Format	0.09	0.08	−0.06	0.23	1.00	10963	6350

Table 2. Fixed effects estimates from Bayesian mixed-effects logistic regression. Dependent variable: Hp choice. Categorical predictors (Duration, Order, Format) were effect-coded, and positive estimates are associated with the level indicated in parentheses. See Methods for details on predictor variables. ESS: Effective sample size. Results obtained from two MCMC chains with 5000 posterior samples and 1000 tuning samples. Terms credibly different from zero are shown in boldface.

difference = −0.4% [−1.6%, 0.8%]; mean $d = -0.05$ [−0.21, 0.10]; $BF_{0+} = 19.34$; Figure 3b).

Strong evidence for effect of presentation order in alternative-wise presentation

We furthermore hypothesized that alternatives shown last (in alternative-wise presentation) and alternatives with better values on the attribute shown last (in attribute-wise presentation) would be chosen more frequently. To test the marginal effect of presentation order on choice (across presentation formats), we computed the individual probabilities of choosing Hp when it was favored by the last presentation stage (because either the alternative Hp or the attribute p was shown last) and when it was not, and performed directed Bayes factor t-tests and Bayesian estimation of their difference. Across presentation formats, participants chose the Hp alternative with a probability of 63.3% when it was favoured in the final presentation stage, and 60.9% when it was not (Table 1). There was extreme evidence in favour of an effect of presentation order on choice across presentation formats (mean difference = 2.4% [1.2%, 3.5%]; mean $d = 0.32$ [0.16, 0.47]; $BF_{+0} = 806.85$). Separate tests for each presentation format showed that the effect was more specific to alternative-wise presentation, with extreme evidence for an effect (H2a; mean difference = 3.0% [1.6%, 4.4%]; mean $d = 0.33$ [0.17, 0.49]; $BF_{+0} = 845.71$; Figure 3c). In attribute-wise presentation, evidence anecdotally favoured a positive effect (H2b; mean difference = 1.5% [0.0%, 3.1%]; mean $d = 0.15$ [−0.01, 0.31]; 96.67% of posterior mass above zero; $BF_{+0} = 1.83$; Figure 3d).

Discussion

In this study, we investigated the causal effects of attribute- and alternative-wise presentation duration and order in two-alternative risky choice. In contrast to causal interpretations of simple gaze-dependent accumulation models, our data did not support a causal role of presentation duration in either presentation format. Instead, we found strong evidence for a causal effect of presentation order on choice, especially when information was presented alternative-wise.

Prior work has reported causal effects of viewing- and presentation duration on preferential (and perceptual) choice (Armel et al., 2008; Fisher, 2021; Liu, Lyu, et al., 2020; Pärnamets et al., 2015; Shimojo et al., 2003; Sui et al., 2020; Tavares et al., 2017) using external control of presentation durations or gaze-contingent decision prompts. In free choice paradigms, these effects are well described by gaze-dependent accumulation models like the aDDM, which assume that an alternative's value representations are discounted while it is not fixated by (or presented to) the decision maker.

In our task, evidence strongly favoured null effects of presentation duration on choice, both in alternative- and attribute-wise presentation formats.

This lack of replication suggests that more work is needed to better understand the conditions under which fixation- or presentation duration can affect choice behaviour. It is possible that duration effects

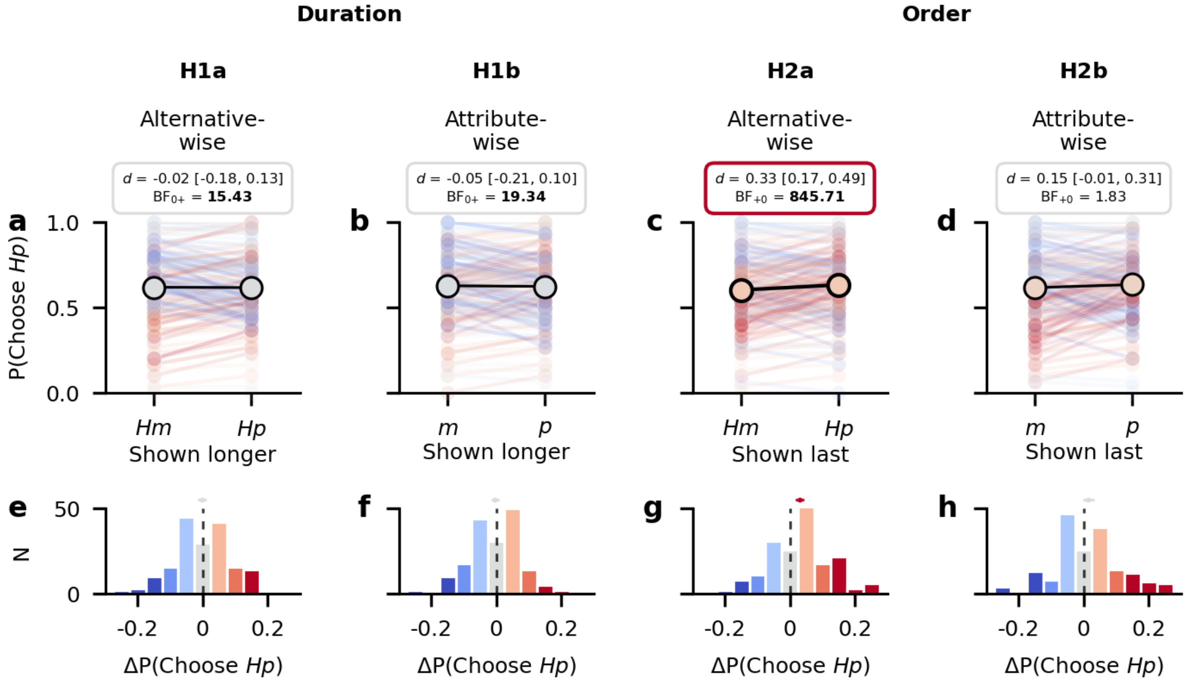


Figure 3. Effects of presentation duration and order on choice. Panels **a-d** show individual and mean changes in choice probabilities of the higher probability alternative Hp, for duration (**a, b**) and order manipulations (**c, d**) in alternative- (**a, c**) and attribute-wise (**b, d**) presentation. Each semi-transparent connected pair of dots indicates choice probabilities for Hp of a single participant in trials where Hm vs. Hp was favoured by the presentation manipulation. Group means are indicated by opaque slate dots and lines. Cohen's d with $\mathrm{HDI}_{95\%}$ obtained from paired BEST analysis, and the Bayes factors in favour of a positive directed over a null effect (BF_{+0}; see Methods) or its reciprocal in favour of a null effect (BF_{0+}) are given for each panel. Panels **e-f** show corresponding distributions of individual changes (individual slopes in panels **a-d**), with colors coding the size of individual changes. Small points and horizontal lines above histograms indicate mean and $\mathrm{HDI}_{95\%}$ change (gray if $\mathrm{HDI}_{95\%}$ includes 0, red if 0 is excluded).

are moderated by specific aspects of the decision task itself, like presentation durations, and the type and delivery of choice stimuli.

It remains, however, an experimental challenge to investigate the causal effects of different aspects of information search on choice: With the external control of presentation parameters, participants are prompted for a decision either before or after they would have made a choice in a free response paradigm. Studies experimentally controlling stimulus presentation (including studies with gaze-dependent decision prompts) thereby differ from simple decision making tasks where participants determine the time of choice themselves; their experimental manipulations interfere with and alter the natural course of decision making. Here, recent work has demonstrated an elegant possibility to induce biases in information search, with downstream effects on choice: Gwinn et al. (2019) had participants acquire attentional biases in a separate task, which carried over to a choice task otherwise free from interference. Notably, the authors found that not gaze duration, but the location of the first fixation, mediated the effect of the attentional manipulation on choice, providing additional evidence of effects of information order on choice over viewing duration alone.

Regarding the association between the order of information search during decision making and choice, most eye tracking studies of decision making find the last fixation to be predominantly directed towards the chosen alternative (Fiedler & Glöckner, 2012; Glickman et al., 2019; Krajbich et al., 2010; Krajbich & Rangel, 2011; Stewart, Hermens, et al., 2016). This association is predicted by gaze-dependent accumulation models like the aDDM, where relative evidence for an alternative is more likely to cross the decision boundary while it is fixated. Notably, however, the aDDM does not imply a causal role of information order or the final fixation on choice in particular. In contrast, leaky accumulator models (e.g., Ashby et al., 2016; Glickman et al., 2019; Molter et al., 2021; Usher & McClelland, 2001) predict that information acquired early in the trial decays, and the relative weight of more recent information is increased.

We found strong evidence for a causal effect of presentation order, so that information presented last influences decisions more than information presented earlier. This recency effect was particularly strong in the alternative-wise presentation format.

Our results thereby support theories of gaze-dependent accumulation which include a form of accumulation leak and can account for causal recency effects. Interestingly, prior work has identified recency effects in decisions from experience and related paradigms, where information about risky alternatives' outcomes is actively sampled by the decision maker, and therefore also acquired sequentially (Hertwig et al., 2004; Tsetsos et al., 2012). Our results suggest that a similar, and causally directed effect is present during decisions from description, where stimulus information is also experienced sequentially.

One possible explanation for the observed recency effect is that decision makers have imperfect memory about the stimulus information (potentially reinforced by the fast-paced presentation) and then base their choice on more recent, better remembered information. This would constitute a memory-bias within single trials, similar to previously described effects on longer timescales (Gluth et al., 2015; Weilbächer et al., 2021).

Our results further provide evidence that the observed association between decision makers' last fixation and choice goes beyond simple "confirmation" or response locking explanations assuming that choices are already determined before the final fixation and response is made.

In addition to the last fixation preferably being directed to the chosen alternative, eye tracking studies of decision making typically find longer gaze towards the chosen alternative (Cavanagh et al., 2014; Fiedler & Glöckner, 2012; Glickman et al., 2019; Krajbich et al., 2010; Krajbich & Rangel, 2011; Molter et al., 2021; Shimojo et al., 2003; Stewart, Hermens, et al., 2016). Both being associated with choice, these two variables might be confounded frequently, that is, alternatives that are looked at last are also looked at longer during the decision. Consequently, the effects of duration and order can be mistaken for each other, without explicit control. This issue can also occur in experimental designs using gaze-dependent decision prompts, as – depending on the specific conditions triggering the prompt – the last fixation is potentially more likely to be directed at the alternative determined to receive longer gaze.

While prior work demonstrated that random or externally induced fluctuations in visual attention can have downstream effects on choice mainly through viewing duration, our work highlights the importance of the temporal order in which information is encountered. This suggests that, in situations where presentation order can be controlled, such as television-, cinema- or online video advertisements, choices could be systematically shifted towards certain alternatives. We note, however, that these settings typically do not involve the need to choose (as in our experiment). Yet, it is conceivable that for those cinema-goers who already decided to buy ice-cream, still deliberating which flavour to get, choices are shifted towards the flavour presented last (and closest to their buying decision) in the advertisement clip. Desired (e.g., healthful) choices could then be promoted by reminding decision makers of them just before choices are made.

Similarly, Sullivan and Huettel (2021) argued in favour of time-dependent interventions to promote healthful food choices, as they found health information to enter the decision process with a longer latency than taste information. Reducing overall time pressure could increase healthful choices by reducing the gap in attribute latencies. Alternatively, they suggest that showing health-related information *first* should decrease the latency gap and thereby promote healthy choices. Our results highlight the opposite direction: Attribute-information shown first to participants reduced participants' likelihood of choosing alternatives with better values on that attribute. Note that our study addressed primacy and recency effects on the same one-dimensional construct. We therefore cannot distinguish positive recency- from negative primacy effects, as information shown last was by design *not* shown first. Future work could address possible independent contributions of both effects.

A possible explanation for these apparent differences, apart from the different goods being chosen, is that the latency construct in the model by Sullivan and Huettel (2021) might be more related to consideration *duration* than *order*: Attributes with lower latencies can, by definition, influence the decision process for a longer duration than others. This way, the order and duration effects are inherently linked in their model.

In conclusion, we showed that presentation order but not presentation duration had a causal effect on risky choice behaviour, in an external presentation paradigm. This has important implications for theories of decision making, which should incorporate mechanisms like accumulation leak, to account for observed recency effects, and suggests an important role of the presentation order – particularly the information presented just before a choice is made – in real world decisions.

Methods

All data collection and analysis procedures were preregistered on the Open Science Framework ahead of data collection (Molter & Mohr, 2021).

Participants

The experiment was conducted online and participants were recruited via the platform Prolific.co. Our target sample size was 200 participants. We accepted entries on the Prolific platform until 200 complete submissions were reached. Sample size was determined to exceed that of previous laboratory-based experiments addressing related questions (e.g., Armel et al., 2008; Fisher, 2021; Liu, Lyu, et al., 2020; Pärnamets et al., 2015; Sui et al., 2020), accounting for the possibility of lower quality data due to the online setting. Participants were paid £3.75 for completing the study, which took around 30 minutes. They had the chance to win a bonus amount ranging from £0 to £9.85, as one of their chosen gambles was randomly determined and played out after the experiment. The experiment included webcam-based eye tracking (see Supplementary Information), so only participants with a working webcam using Google Chrome or Mozilla Firefox could participate. Participants gave informed consent prior to participation in the study.

Participants were excluded from the analysis without replacement if one of the following criteria was met: i) The participant chose a dominated alternative more than 4 times (20% of catch trials, see below). ii) The participant reported to be red-green colorblind or having difficulty distinguishing the colors in the task. iii) The participant reported having technical difficulties that prevented them from diligent execution of the task. iv) The participant's self-reported decision strategy suggested that task instructions were misunderstood. v) The participant reported nonserious participation in the task (Aust et al., 2013).

Twenty-one participants were excluded from the analyses (16 for criterion 1, four for criterion 2, one for criterion 4), resulting in a final sample size of 179 (mean ± s.d. age = 32.46 ± 13.85; 84 females, 94 males, one other).

Behavioral task

Participants performed a binary sequential presentation risky choice task, where they made repeated incentivized choices between two graphically displayed all-or-nothing gambles (Figure 2). Each gamble represented the chance to win a monetary amount m with a probability p and nothing otherwise. Winning amounts m were represented by partially filled bars (a fully filled bar represented an amount of 10£), and winning probabilities p by pie charts. Each trial was divided into a presentation phase and a choice phase. During the presentation phase, participants learned about the two available gambles' attributes. Information was presented sequentially. There were two types of presentation: Alternative-wise and attribute-wise presentation.

In trials with alternative-wise presentation (Figure 2a) both attributes of one gamble were shown simultaneously, followed by both attributes of the other gamble. In trials with attribute-wise presentation (Figure 2b), one attribute (e.g., winning probability p) of both gambles was presented simultaneously, followed by the other attribute (e.g., amount m) of both gambles. Presentation always alternated two times between alternatives or attributes (i.e., A-B-A-B). Crucially, one alternative (in trials with alternative-wise presentation) or attribute (in trials with attribute-wise presentation) in each trial was selected to be the target and shown longer than the other one. Target attributes or alternatives were always presented for 1500 ms, whereas other alternatives or attributes were presented for 1000 ms, resulting in a final presentation time advantage for the target of 1000 ms (2x 1500 ms vs. 2x 1000 ms).

After the presentation phase, stimulus information was hidden and participants were prompted to make a choice between the two alternatives within 3 seconds. After completing all choice trials, one gamble chosen by the participant in a randomly determined trial was played out for a real bonus payment. The task can be run at https://moltaire.github.io/causality_task.

All participants made choices for the same 140 choice problems (see below) divided into two blocks. Trial order in each block and the horizontal position of the alternatives in each trial was randomized. Block order was counterbalanced between participants.

Stimuli

We created a set of 15 core choice problems including one alternative with a higher probability of winning a lower amount (Hp) and one alternative with a lower probability of winning a higher amount (Hm). Choice problems were created algorithmically to cover most of the attribute space, be maximally different from each other, and be diagnostic of different risk attitudes. For this last criterion, we controlled the α values for which two alternatives in a pair would have equal expected utility (using a standard power utility function). The distribution of these indifference-implied α values is shown in in Figure 2d. Core choice problems are illustrated in Figure 2c.

Then eight trials were created for each core problem by fully crossing the factors (i) presentation format (alternative-wise vs. attribute-wise), (ii) target alternative / attribute (Hp vs. Hm; p vs. m), and (iii) presentation order (target first and third vs. second and last). This resulted in a total of 120 experimental choice trials.

We added 20 catch trials with one dominant alternative for a total of 140 trials (Figure 2e). In catch trials, individual presentation durations were set to 1250 ms, resulting in the same overall presentation duration (5000 ms), but no presentation time advantage for any alternative or attribute.

Statistical modelling

We performed a Bayesian logistic regression analysis of choice behavior, with choice (Hp vs. Hm) as the dependent variable and the following predictors: Expected value difference ($EV_{Hp} - EV_{Hm}$; z-scored), presentation format (by-attribute vs. by-alternative; effect-coded), duration-favored (Hp favored vs. Hm favored by duration manipulation; effect-coded), last-stage-favored (Hp favored vs. Hm favored in last presentation stage; effect-coded), and interaction terms between presentation format and duration-favored and last-stage-favored. The model included random intercepts and slopes over participants and used the default priors set by the bambi library (Westfall, 2017).

We performed directed Bayes factor (BF) t-tests (Morey & Rouder, 2011) of our main hypotheses, testing the directed hypotheses that differences in choice probabilities are larger than zero, over a point null hypothesis. All Bayes factor t-tests used default JZS priors (Cauchy distributed with scale $r = \sqrt{2}/2$) implemented in the BayesFactor package (Morey & Rouder, 2018).

Additionally, we ran paired Bayesian estimation (BEST; Kruschke, 2013, 2014) analyses to compute mean differences, effect sizes d and associated 95% highest posterior density intervals ($\mathrm{HDI}_{95\%}$).

We ran two chains with 5000 samples each after a tuning phase of 1000 samples for all BEST analyses and the Bayesian mixed-effects model. Convergence was diagnosed visually and by means of the Gelman-Rubin statistic ($|1 - \hat{R}| \leq 0.05$ for all chains).

We determine parameters in the regression and BEST analyses to be credibly different from zero if $\mathrm{HDI}_{95\%}$ exclude zero or at least 95% of the posterior mass is above (below) zero. For interpretation of Bayes factors' evidence strength, we follow the conventional categorization based on Jeffreys (1998): Anecdotal ($1 < \mathrm{BF} < 3$); Moderate ($3 \leq \mathrm{BF} < 10$); Strong ($10 \leq \mathrm{BF} < 30$); Very strong ($30 \leq \mathrm{BF} < 100$); Extreme ($100 \leq \mathrm{BF}$).

Software

The task was programmed in jsPsych (de Leeuw, 2015). Webcam-based eye tracking was implemented using the webgazer.js library (Papoutsaki et al., 2016). Data processing and analyses were done in Python with numpy (Harris et al., 2020) and pandas (McKinney, 2012) libraries. Bayesian analyses were implemented in PyMC3 (Salvatier et al., 2016), mixed models used bambi (Capretto et al., 2021). Bayes Factor t-tests were performed using the R package BayesFactor (Morey & Rouder, 2018). Figures were created using matplotlib (Hunter, 2007).

Data and code availability

All raw and preprocessed data and scripts to reproduce all processing and analyses steps and figures are available at https://github.com/moltaire/gaze-choice-causality.

References

Allais, M. (1953). L'Extension des Theories de l'Equilibre Economique General et du Rendement Social au Cas du Risque. *Econometrica*, *21*(2), 269–290. https://doi.org/10.2307/1905539

Amasino, D. R., Sullivan, N. J., Kranton, R. E., & Huettel, S. A. (2019). Amount and time exert independent influences on intertemporal choice. *Nature Human Behaviour*, *3*(4), 383–392. https://doi.org/10.1038/s41562-019-0537-2

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, *3*(5), 8.

Ashby, N. J. S., Jekel, M., Dickert, S., & Glöckner, A. (2016). Finding the right fit: A comparison of process assumptions underlying popular drift-diffusion models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(12), 1982–1993. https://doi.org/10.1037/xlm0000279

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, *45*(2), 527–535. https://doi.org/10.3758/s13428-012-0265-2

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459. https://doi.org/10.1037/0033-295X.100.3.432

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2021). Bambi: A simple interface for fitting Bayesian linear models in Python. *arXiv:2012.10754 [stat]*.

Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology. General*, *143*(4), 1476–1488. https://doi.org/10.1037/a0035813

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Fiedler, S., & Glöckner, A. (2012). The Dynamics of Decision Making in Risky Choice: An Eye-Tracking Analysis. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00335

Fisher, G. (2021). Intertemporal Choices Are Causally Influenced by Fluctuations in Visual Attention. *Management Science*. https://doi.org/10.1287/mnsc.2020.3732

Glickman, M., Sharoni, O., Levy, D. J., Niebur, E., Stuphorn, V., & Usher, M. (2019). The formation of preference in risky choice. *PLOS Computational Biology*, *15*(8), e1007201. https://doi.org/10.1371/journal.pcbi.1007201

Glöckner, A., & Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*(1), 71–98. https://doi.org/10.1002/bdm.684

Gluth, S., Kern, N., Kortmann, M., & Vitali, C. L. (2020). Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nature Human Behaviour*, 1–12. https://doi.org/10.1038/s41562-020-0822-0

Gluth, S., Sommer, T., Rieskamp, J., & Büchel, C. (2015). Effective connectivity between hippocampus and ventromedial prefrontal cortex controls preferential choices from memory [Publisher: Elsevier]. *Neuron*, *86*(4), 1078–1090. https://doi.org/10.1016/j.neuron.2015.04.023

Gluth, S., Spektor, M. S., & Rieskamp, J. (2018). Value-based attentional capture affects multi-alternative decision making (M. J. Frank, Ed.). *eLife*, *7*, e39659. https://doi.org/10.7554/eLife.39659

Gwinn, R., Leber, A. B., & Krajbich, I. (2019). The spillover effects of attentional learning on value-based choice. *Cognition*, *182*, 294–306. https://doi.org/10.1016/j.cognition.2018.10.012

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del R'ıo, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from Experience and the Effect of Rare Events in Risky Choice. *Psychological Science*, *15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, & Psychophysics*, *81*(7), 2288–2303. https://doi.org/10.3758/s13414-019-01846-w

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Isham, E. A., & Geng, J. J. (2013). Looking Time Predicts Choice but Not Aesthetic Value. *PLOS ONE*, *8*(8), e71698. https://doi.org/10.1371/journal.pone.0071698

Jeffreys, S. H. (1998). *The Theory of Probability* (Third Edition). Oxford University Press.

Kahneman, D., & Tversky, A. (2012). Choices, Values, and Frames. *Handbook of the Fundamentals of Financial Decision Making* (pp. 269–278). World Scientific. https://doi.org/10.1142/9789814417358_0016

Kim, B., Seligman, D., & Kable, J. W. (2012). Preference Reversals in Decision Making Under Risk are Accompanied by Changes in Attention to Different Attributes. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00109

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298. https://doi.org/10.1038/nn.2635

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, *3*, 193. https://doi.org/10.3389/fpsyg.2012.00193

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13852–13857. https://doi.org/10.1073/pnas.1101328108

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573–603. https://doi.org/10.1037/a0029146

Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press
OCLC: ocn861318341.

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*(37), 13214–13223. https://doi.org/10.1523/JNEUROSCI.1246-11.2011

Liu, H.-Z., Lyu, X.-K., Wei, Z.-H., Mo, W.-L., Luo, J.-R., & Su, X.-Y. (2020). Exploiting the dynamics of eye gaze to bias intertemporal choice. *Journal of Behavioral Decision Making*, *n/a*(n/a). https://doi.org/10.1002/bdm.2219

Liu, H.-Z., Zhou, Y.-B., Wei, Z.-H., & Jiang, C.-M. (2020). The power of last fixation: Biasing simple choices by gaze-contingent manipulation. *Acta Psychologica*, *208*, 103106. https://doi.org/10.1016/j.actpsy.2020.103106

Lopez-Persem, A., Domenech, P., & Pessiglione, M. (2016). How prior preferences determine decision-making frames and biases in the human brain. *eLife*, *5*. https://doi.org/10.7554/eLife.20317

Maier, S. U., Raja Beharelle, A., Polanía, R., Ruff, C. C., & Hare, T. A. (2020). Dissociable mechanisms govern when and how strongly reward attributes affect decisions. *Nature Human Behaviour*, 1–15. https://doi.org/10.1038/s41562-020-0893-y

McKinney, W. (2012). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. "O'Reilly Media, Inc."

Milosavljevic, M., Navalpakkam, V., Koch, C., & Rangel, A. (2012). Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology*, *22*(1), 67–74. https://doi.org/10.1016/j.jcps.2011.10.002

Mittone, L., & Papi, M. (2020). Inducing alternative-based and characteristic-based search procedures in risky choice. *Judgment and Decision Making*, 10.

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*(1), 1–10. https://doi.org/10.1038/s41598-017-06968-5

Molter, F., & Mohr, P. N. C. (2021). Investigating the causal effects of presentation duration, sequence, and format on binary risky choice. *Open Science Framework*. https://osf.io/bth3m

Molter, F., Thomas, A., Huettel, S. A., Heekeren, H., & Mohr, P. N. C. (2021). Gaze-dependent evidence accumulation predicts multi-alternative risky choice behaviour. https://doi.org/10.31234/osf.io/x6nbf

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*(4), 406–419. https://doi.org/10.1037/a0024377

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of bayes factors for common designs*. Manual.

Mormann, M., & Russo, J. E. (2021). Does Attention Increase the Value of Choice Alternatives? *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2021.01.004

Mullett, T. L., & Stewart, N. (2016). Implications of visual attention phenomena for models of preferential choice. *Decision*, *3*(4), 231. https://doi.org/10.1037/dec0000049

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. *Pro-*

ceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI), 3839–3845.

Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., & Richardson, D. C. (2015). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(13), 4170–4175. https://doi.org/10.1073/pnas.1415250112

Reeck, C., Wall, D., & Johnson, E. J. (2017). Search predicts and changes patience in intertemporal choice. *Proceedings of the National Academy of Sciences*, *114*(45), 11890–11895. https://doi.org/10.1073/pnas.1707040114

Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies — Revisited. *NeuroImage*, *84*, 971–985. https://doi.org/10.1016/j.neuroimage.2013.08.065

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, *2*, e55. https://doi.org/10.7717/peerj-cs.55

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Sepulveda, P., Usher, M., Davies, N., Benson, A. A., Ortoleva, P., & De Martino, B. (2020). Visual attention modulates the integration of goal-relevant evidence and not value (C. Büchel, V. Wyart, A. L. Filipowicz, & S. Gluth, Eds.). *eLife*, *9*, e60705. https://doi.org/10.7554/eLife.60705

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322. https://doi.org/10.1038/nn1150

Shinn, M., Lam, N. H., & Murray, J. D. (2020). A flexible framework for simulating and fitting generalized drift-diffusion models (T. Kahnt, J. I. Gold, L. Ding, & E.-J. Wagenmakers, Eds.). *eLife*, *9*, e56938. https://doi.org/10.7554/eLife.56938

Smith, S. M., & Krajbich, I. (2018). Attention and choice across domains. *Journal of Experimental Psychology: General*, *147*(12), 1810–1826. https://doi.org/10.1037/xge0000482

Smith, S. M., & Krajbich, I. (2019). Gaze Amplifies Value in Decision Making. *Psychological Science*, *30*(1), 116–128. https://doi.org/10.1177/0956797618810521

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Stewart, N., Gächter, S., Noguchi, T., & Mullett, T. L. (2016). Eye Movements in Strategic Choice. *Journal of Behavioral Decision Making*, *29*(2-3), 137–156. https://doi.org/10.1002/bdm.1901

Stewart, N., Hermens, F., & Matthews, W. J. (2016). Eye Movements in Risky Choice: Eye Movements in Risky Choice. *Journal of Behavioral Decision Making*, *29*(2-3), 116–136. https://doi.org/10.1002/bdm.1854

Sui, X.-Y., Liu, H.-Z., & Rao, L.-L. (2020). The timing of gaze-contingent decision prompts influences risky choice. *Cognition*, *195*, 104077. https://doi.org/10.1016/j.cognition.2019.104077

Sullivan, N. J., & Huettel, S. A. (2021). Healthful choices depend on the latency and rate of information accumulation. *Nature Human Behaviour*, 1–9. https://doi.org/10.1038/s41562-021-01154-0
Bandiera_abtest: a Cg_type: Nature Research Journals Primary_atype: Research Subject_term: Computational neuroscience;Human behaviour Subject_term_id: computational-neuroscience;human-behaviour

Tavares, G., Perona, P., & Rangel, A. (2017). The Attentional Drift Diffusion Model of Simple Perceptual Decision-Making. *Frontiers in Neuroscience*, *11*, 468. https://doi.org/10.3389/fnins.2017.00468

Thomas, A. W., Molter, F., & Krajbich, I. (2021). Uncovering the computational mechanisms underlying many-alternative choice (T. Verstynen, M. J. Frank, & E. Peterson, Eds.). *eLife*, *10*, e57012. https://doi.org/10.7554/eLife.57012

Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R., & Mohr, P. N. C. (2019). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, *3*(6), 625–635. https://doi.org/10.1038/s41562-019-0584-8

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*(24), 9659–9664. https://doi.org/10.1073/pnas.1119569109

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. https://doi.org/10.1037/0033-295X.108.3.550

Weber, E., & Kirsner, B. (1997). Reasons for Rank-Dependent Utility Evaluation. *Journal of Risk and Uncertainty*, *14*(1), 41–61. https://doi.org/10.1023/A:1007769703493

Weilbächer, R. A., Krajbich, I., Rieskamp, J., & Gluth, S. (2021). The influence of visual attention on memory-based preferential choice. *Cognition*, *215*, 104804. https://doi.org/10.1016/j.cognition.2021.104804

Westbrook, A., van den Bosch, R., Määttä, J. I., Hofmans, L., Papadopetraki, D., Cools, R., & Frank, M. J. (2020). Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. *Science*, *367*(6484), 1362–1366. https://doi.org/10.1126/science.aaz5891

Westfall, J. (2017). Statistical details of the default priors in the bambi library. *arXiv:1702.01201 [stat]*. Retrieved February 3, 2020, from http://arxiv.org/abs/1702.01201

Yang, X., & Krajbich, I. (2020). Webcam-based online eye-tracking for behavioral research. https://doi.org/10.31234/osf.io/qhme6

Yechiam, E., & Hochman, G. (2013). Loss-aversion or loss-attention: The impact of losses on cognitive performance. *Cognitive Psychology*, *66*(2), 212–231. https://doi.org/10.1016/j.cogpsych.2012.12.001

Author contributions

Conceptualization: F.M. and P.N.C.M. Data curation: F.M. Formal analysis: F.M. Funding acquisition: P.N.C.M. Investigation: F.M. Methodology: F.M. Project administration: P.N.C.M. Resources: F.M. Software: F.M. Supervision: P.N.C.M. Validation: F.M. Visualization: F.M. Writing - original draft: F.M. Writing - review & editing: F.M. and P.N.C.M.

Competing Interests Statement

The authors declare no competing interests.

Supplementary Information

Webcam-based eye tracking

Participants' gaze during the task was recorded using webcam-based eye tracking implemented in the JavaScript library webgazer.js (Papoutsaki et al., 2016; Yang & Krajbich, 2020). To this end, the task included a webcam setup at the start of the experiment, a 13-point calibration routine, and a 5-point validation routine before the start of each block, where participants were instructed to focus their gaze on indicated screen locations for three seconds. Validation points were chosen to correspond to the locations where stimulus information would be presented during experimental choice trials.

We assessed the quality of the collected eye tracking data using multiple complementing metrics: First, we measured the sampling rates obtained in each validation, which heavily depend on the device that runs the experimental task. Mean $\pm$ s.d. sampling rates were 15.96 $\pm$ 7.95 Hz and ranged from 0.49 Hz to 29.55 Hz, and 71% of blocks had sampling rates above 10 Hz during validation.

Second, as a measure of bias, we computed the average absolute distance of the estimated gaze location samples from the corresponding validation targets in x- and y directions (in percent of screen width and height, respectively). Mean $\pm$ s.d. error (clustered by block) was 5.58 $\pm$ 7.11% in x- and 8.79 $\pm$ 9.76% in y-direction.

Third, as a measure of accuracy, we computed the sample variability as the standard deviation of samples corresponding to a single validation target in x- and y-directions. Mean $\pm$ s.d. variability was 5.81 $\pm$ 3.07% in x- and 7.46 $\pm$ 4.70% in y-direction.

Fourth, we computed the proportion of samples within an elliptical Area of Interest (AoI) around the corresponding validation target. For this, we set the AoI-width and -height to 15% of the screen width and height, respectively (note that widths and heights of 25% would leave no white space between AoIs due to the stimulus spacing used in the task). On average, only 51.80 $\pm$ 26.12% of the samples were contained within a validation target's AoI.

Next, we set minimum thresholds on each of these parameters to classify validation results as valid or invalid: We required a minimum sampling rate of 10 Hz, a maximum average error and a maximum variability of 20% in x- and y-directions, and a minimum of 60% of samples within the validation target AoI. Only 33 of 356 blocks (9.27%) fulfilled these minimal quality requirements. In addition, it can be assumed that eye tracking quality only *decreased* during a block, due to head movement or other sources of variability (e.g., changes in lighting conditions, etc.). We therefore did not perform any further explorative analyses on the eye tracking data. We note that it might be possible to correct for systematic biases in validation (e.g., using clustering approaches). Future studies could benefit from improved webcam-based eye tracking data by performing online checks of validation accuracy and repeating calibration and validation steps until an acceptable level of quality is reached.

Influence of presentation format on value integration

Our preregistration contained the additional hypothesis that presentation format (alternative- vs. attribute-wise presentation) elicits different forms of value integration in the decision process. Specifically, we hypothesized, that alternative-wise presentation would elicit more integrative, within-alternative processing, whereas attribute-wise presentation would be associated with more comparative processing within attribute dimensions and between alternatives. To test this hypothesis, we fit two behavioral models, which used within-alternative and within-attribute integration of attributes, respectively, to the choice and response time data of each participant. Crucially, the models were fit separately for trials with alternative- and attribute-wise presentation. We then computed the relative fit of the models for each participant and presentation format by taking the difference of the models' Bayesian Information Criterion (BIC) (Schwarz, 1978). Finally, we performed directed paired Bayes factor t-tests of the differences, testing the directed hypothesis that the relative fit of the within-alternative integration model over the between-alternative model is increased in alternative-wise vs. attribute-wise presentation.

Behavioural modelling

We analyzed participants' choice and response time data using two different behavioral models. Both models shared a similar general structure: During the presentation phase of the trial, a relative evidence signal R is assumed to be formed (Figure S1a-b), depending on the presentation format, duration, and order. After the choice prompt, a noisy diffusion process between two decision bounds (corresponding to choosing Hp and Hm, respectively) is initiated that elicits a choice at a time point t (Figure S1c). Crucially, the drift rate of the diffusion process is proportional to the relative evidence signal R at the end of the presentation phase. For both models, the diffusion process after the choice prompt is parameterized by a drift constant v (that scales the final relative evidence signal R), a noise parameter s, while the boundary separation is kept constant at a value of 1. The two models only differed in the process of calculating the relative evidence signal R during the presentation phase.

Alternative-wise integration model The alternative-wise integration model (slate-gray in Figure S1) computes alternative-wise expected utilities, using a standard utility function ($U_i = p_i m_i^{\alpha}$). During the presentation phase, the difference between the expected utilities is assumed to accumulate over time. Critically, the momentarily not presented alternative's utility is discounted by an alternative-wise gaze-discount θ. The final relative evidence signal $R_{altwise}$ is given by

$$R_{altwise} = g_{Hp}(U_{Hp} - \theta U_{Hm}) + g_{Hm}\theta U_{Hp} - U_{Hm})$$

Where g_{Hp} and g_{Hm} are the relative presentation durations of Hp and Hm. Note that in trials with attribute-wise presentation, g_{Hp} and g_{Hm} are set to 0.5, as both alternatives' attributes are presented equally long. Additionally, the model uses a parameter b_{last} to predict order effects in trials with alternative-wise presentation: Positive b_{last} shift R towards the last-presented alternative, negative b_{last} shift it to the alternative presented first.

Attribute-wise integration model The attribute-wise model (orange in Figure S1) assumes that the relative evidence R is computed through attribute comparisons between alternatives, weighted addition of attribute differences, and accumulation of differences over time. Importantly, it assumes that the momentarily not presented attributes are discounted by an attribute-wise gaze-discount η. The final relative evidence signal $R_{attwise}$ is given by

$$\begin{aligned} R_{attwise} = &g_p(w_p\Delta_p + \eta(1 - w_p)\Delta_m) + \\ &g_m(\eta w_p\Delta_p + (1 - w_p)\Delta_m) \end{aligned}$$

Where Δ_p and Δ_m are the attribute differences between Hp and Hm alternatives. g_p and g_m are relative presentation durations for p and m attributes, respectively. w_p controls the relative weighting between probability and outcome attributes. Note that attributes are also normalized in each trial (by dividing by the sum of values on the attribute). During trials with alternative-wise presentation, the relative gaze durations towards attributes are set to 0.5, since at every point during the presentation phase, information of both attributes is presented. Additionally, the model uses a parameter b_{last} to predict order effects in trials with attribute-wise presentation: Positive b_{last} shift R towards the alternative with the higher value on the last-presented attribute, negative b_{last} shift it to the alternative with the higher value on the alternative presented first.

Parameter estimation Both models were implemented in pyddm (Shinn et al., 2020) with a temporal resolution $dt = 0.01$, and a resolution of the evidence space $dx = 0.01$. Both models were fit separately to trials with alternative-wise and attribute-wise presentation using pyddm's default differential evolution algorithm, minimizing the Bayesian Information Criterion (BIC; Schwarz, 1978).

Model validation To ensure interpretability and validity of the models' parameter estimates, we performed a parameter recovery study as follows: First, we estimated each model's parameters from the empirical data of each participant. Then we simulated a synthetic data set of the same size as the empirical one, using the individually obtained estimates. Then we re-fit the models to the synthetic data and compared known generating to the obtained recovered parameters by means of Bayesian linear regression (dependent variable: recovered parameter; independent variables: Intercept, generating parameter) and correlation analyses (Kruschke, 2013; Lee & Wagenmakers, 2013). The models' parameters could be recovered to a satisfying degree, with the gaze-discount parameters θ and η showing the largest differences (Figure S2).

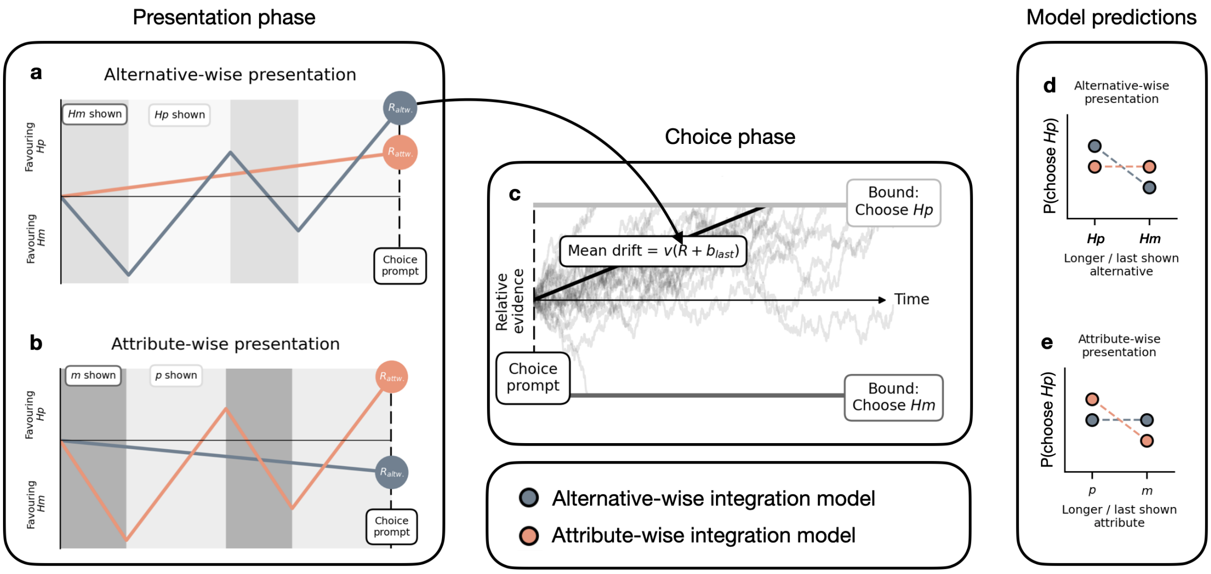


Figure S1. Behavioural models and their predictions. a) Construction of drift-rates in trials with alternative-wise presentation. Only the alternative-wise model's drift rate is sensitive to effects of presentation duration in alternative-wise presentation. Analogously, only the alternative-wise model produces order effects in alternative-wise presentation, controlled by the b_{last} parameter. **b)** Construction of drift rates in attribute-wise presentation. Here, only the attribute-wise model's drift rate is sensitive to effects of presentation duration. Similarly, only the attribute-wise model produces order effects in attribute-wise presentation, controlled by the b_{last} parameter. **c)** After construction of the drift rate in the presentation phase, choices and response times result from a diffusion process with the previously constructed drift rate. **d-e)** Both models' predicted effects of alternative-wise (**d**) and attribute-wise (**e**) presentation duration. Slate-gray color refers to alternative-wise model. Orange color refers to attribute-wise model.

Similarly, we performed model recovery analyses by fitting each model to the synthetic data generated from all models. Then, for each generating model, we performed Bayesian model selection (Rigoux et al., 2014; Stephan et al., 2009) to identify the most likely generating model. The models could be recovered almost perfectly (Figure S3).

Results

For trials with alternative-wise presentation, the mean $\pm$ s.d. BIC of the alternative-wise model was 102.16 $\pm$ 56.19 (range -24.20 to 344.44). Mean $\pm$ s.d. BIC of the attribute-wise model was 93.27 $\pm$ 57.03 (range -45.06 to 338.90), indicating a better fit of the attribute-wise model in alternative-wise presentation.

Conversely, in trials with attribute-wise presentation, the mean $\pm$ s.d. BIC of the alternative-wise model was 90.97 $\pm$ 55.24 (range -45.39 to 310.33), while the attribute-wise model achieved mean $\pm$ s.d. BIC of 97.41 $\pm$ 55.36 (range -50.91 to 321.62), indicating better fit of the alternative-wise model.

Next, we took the difference between BIC of the models for each participant and presentation format and tested whether they credibly differed from zero, using a directed paired Bayes factor t-test (and BEST analysis to obtain estimates of the effect size d). Results strongly supported a result *opposite* to our hypothesis, namely, that the attribute-wise model was favoured in alternative-wise presentation, while the alternative-wise model was favoured in attribute-wise presentation (mean BIC difference = -15.34 [-14.57, -16.20]; mean d = 2.75 [2.29, 3.23]; BF_{+0} = 0.0011, BF_{0+} = 907.95).

In sum, these results are contrary to our preregistered hypothesis: In trials with alternative-wise presentation, the simpler attribute-wise model was preferred. Vice versa, in trials with attribute-wise presentation, the simpler alternative-wise model was preferred.

We note, however, that the preregistered analysis could not address our hypothesis optimally for multiple reasons. First, model complexity differed systematically between conditions: In trials with alternative-

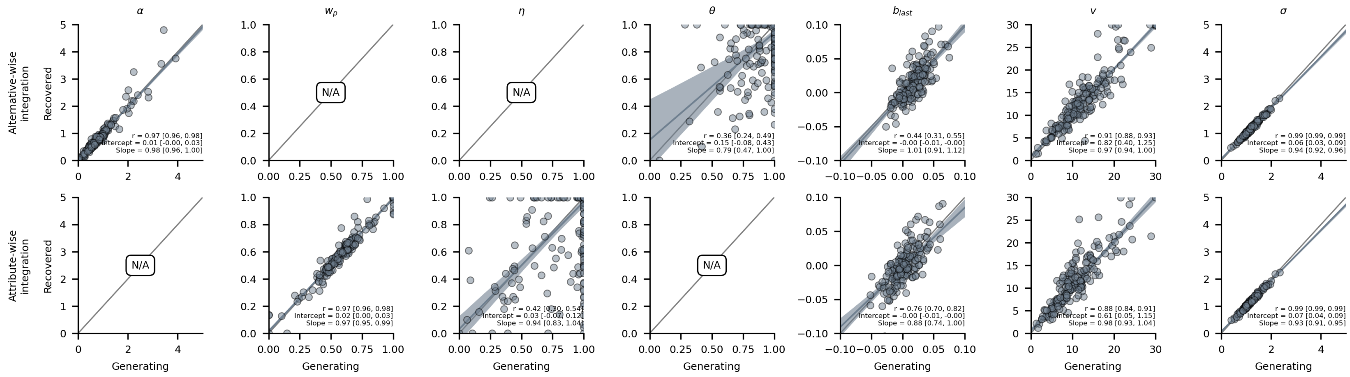


Figure S2. Parameter recovery of the two diffusion models. Each panel shows the relationship between true data generating, and recovered parameter values for a single model parameter. Upper and lower rows show results from the *alternative-wise* and *attribute-wise* integration models, respectively. Generating parameters were obtained from fitting the models to the empirical data. Annotations report Bayesian correlation coefficient r, and the slope and intercept estimates of a Bayesian regression analysis with $\text{HDI}_{95\%}$ given in brackets. Perfect, unbiased recovery would show an intercept of 0 and a slope of 1, with all points on the diagonal.

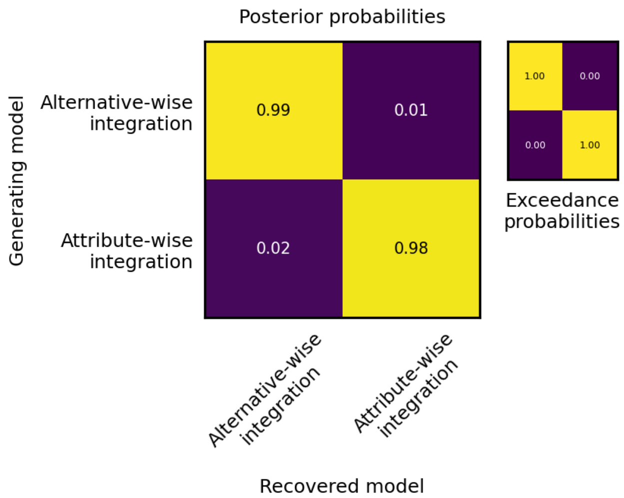


Figure S3. Model recovery results. The large panel shows a confusion matrix from the model recovery analysis. Each cell shows the posterior model probability of a fitted model (in each column) for a given generating model (in each row). Perfect recovery would show only values of 1 on the diagonal. The smaller confusion matrix shows exceedance probabilities, analogously.

wise presentation, the alternative-wise model uses two free parameters more than the attribute-wise model (namely, the alternative-wise gaze discount θ, and the alternative-wise b_{last} parameter), while the reverse is true in trials with attribute-wise presentation. While the BIC takes model complexity into account, this issue highlights that the models differ in more aspects than their value-integration process (namely, the different gaze discounts and last-stage effects), prohibiting conclusions about changes in value-integration. Furthermore, more complex models are penalized for their inclusion of duration-dependent gaze-discount mechanisms to explain associations of presentation duration and choice, which our behavioural analyses showed to be absent in our data.