



OPEN Probabilistic functionalism as a limiting condition for robustness

Mattias Forsgren¹✉, Benjamin Mandl^{2,3} & Gustav Karreskog Reh binder^{4,5}

When should we expect behavioural phenomena to be robust? We argue that many phenomena of interest to behavioural scientists, by their very nature, involve manipulations of stimulus characteristics. If there exist contingencies between those stimulus characteristics and outcomes, the former will consequently constitute cues. People may then pick up on whether the cue is guiding them towards their goal or not and adapt their behaviour accordingly. On this view, the robustness of such phenomena is, at least partly, determined by the cue structure in each given setting. In an experiment, we demonstrate that the attraction effect and the default nudge obtain proportionally to how well the manipulated stimulus characteristics predict the superior option. A similar result is found under a more traditional rule learning manipulation. We suggest that the existence of cue-outcome relationships is an (underappreciated) limiting condition for the robustness of behavioural phenomena. We discuss the implications of this perspective for nudging.

Keywords Probabilistic functionalism, Robustness, Attraction effect, Default nudge

We may consider a behavioural phenomenon to be robust if conceptually similar experimental manipulations consistently produce effects on behaviour in the same direction and of comparable size. High robustness is central to discussions of external validity¹, theory testing² and the replication crisis^{3–5}. Discussions on the limits of robustness have, we believe, tended to take a statistical perspective (e.g.^{6–8}). Here, we suggest a theoretically motivated limiting condition for robustness and demonstrate it with empirical examples. While it cannot be applied to all behavioural phenomena, we propose that it has relevance for the rather broad class of studies where the experimental manipulation consists of adapting stimulus characteristics or the context in which they are presented.

A natural but, it seems to us, overlooked perspective on robustness can be found in Brunswikian^{9,10} psychology. The “lens model” (Fig. 1) distinguishes between distal, unobserved properties and proximal, observed cues. In any given setting (“ecology”), there exists statistical relationships between the cues and properties. Research in the Brunswikian tradition supposes that the mind is geared towards calibrating its judgement strategies such that it makes the most effective use of the cues to predict the properties (“probabilistic functionalism”,¹¹). Cues are, however, typically only imperfect predictors of properties and the challenge for the organism is to learn how much weight to put on each cue. The best one can do is to adapt one’s weighting of cues such that it is equal to how strongly a cue actually predicts a property (i.e. calibrate “cue weights” to be equal to “cue predictivities”,¹²). People appear to be decent at this but performance depends on characteristics of the task and feedback (see¹³ for a review). When cues have a simple (e.g. linear) relationship to the outcome variable (e.g.¹⁴), this learning is believed to be declarative^{15,16}; the extracted contingency can be verbalised, allowing participants to extrapolate beyond the observed cue values.

We tentatively suggest that a theoretical commitment to probabilistic functionalism should make us qualify our expectations of robustness. When experiments involve manipulating characteristics of the options, these manipulations create regularities in the environment that constitute cues. For example, if we manipulate the set of options presented (e.g.¹⁷), or their attribute values (e.g.¹⁸), then which kinds of options that co-occur can constitute a cue. If we manipulate the user interface layout (e.g.¹⁹), then how options are ordered, where they are located on the screen, or which option is pre-selected can constitute a cue. If we manipulate visual characteristics of individual options (e.g.²⁰), their shape, colour, or pattern can constitute cues. Participants may then pick up the cue-outcome relationship and adapt their behaviour accordingly. We should thus expect robustness in settings where no such relationships exist, are inferior to some other policy of choice, or where the conditions are too severe to allow adaptation of cue weights. Further downstream, if we believe that phenomena will be robust only in lieu of predictive cue-outcome relationships, then this constitutes a limiting condition on (i) the external

¹Department of Psychology, Uppsala University, P. O. Box 1225, 751 42 Uppsala, Sweden. ²Department of Economics, Stockholm School of Economics, Stockholm, Sweden. ³Verbund AG, Vienna, Austria. ⁴Institute for Futures Studies, Stockholm, Sweden. ⁵Department of Economics, Uppsala University, Uppsala, Sweden. ✉email: mattias.forsgren@psyk.uu.se

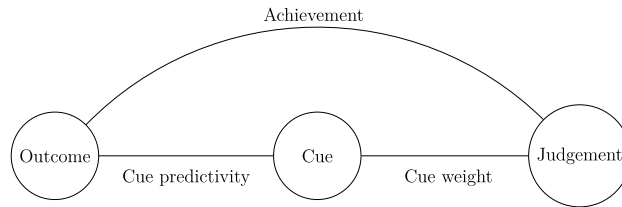


Fig. 1. A simplified lens model adapted to the current experiment. There exists some objective association between the distal Outcome and proximal Cue (Cue predictivity). The judge may extract this relationship and adapt their judgement to it (Cue weight). Achievement is how well the judgement corresponds to the outcome.

validity and replicability⁴ of phenomena and (ii) the scope of the theories they motivate. We tentatively suggest that researchers sometimes over-generalise their findings when they neglect this limitation (e.g.^{21,22}).

Apart from this Brunswikian perspective, with its roots in perception psychology, similar thinking has been put forward in social psychology. (We are grateful to an anonymous reviewer for reminding us of this). For example, rather than “cues” and “cue weights”, Schwarz and Bless²³ speak of “sources of information” providing “routes to judgement”, and “filters” that determine which sources are utilised. Wegener and Petty²⁴ emphasised the metacognitive nature of filtering, seeing it as driven by the individual’s beliefs about how different sources of information exerted (undesired) influence on their initial judgement (cf.²⁵). This filtering will only be well-tuned in so far as people’s beliefs are accurate and they are motivated and able to filter (see e.g.²⁶). This is perhaps the clearest contrast to probabilistic functionalists who instead tend to assume that cue weights will (eventually) be well-calibrated (e.g.²⁷) — at least to the extent allowed by the constraints imposed by the environment (cf.²⁸). So, while we do not deny that there are nuanced theoretical differences between these traditions, they share the fundamental idea that cue use is adaptive. Despite this recurrent evolution, “this point seems to be strangely both intuitive yet subtle, such that it needs to keep being rediscovered, especially in different subfields” (as phrased by²⁹, albeit regarding a different issue).

We do not believe that the perspective described above can be “proven” in any direct sense, much like we do not believe that one can prove that experimental results are, or are not, externally valid *in general* (cf.^{30,31}). What one can do, however, is to build a library of cases where it indeed turned out that cue predictivity was a limiting condition for robustness (cf.^{32,33} in the context of external validity). Such demonstrations are of direct relevance to the scrutinised cases, but might also motivate a more general notion of the perspective being important. We submit a contribution to that library here by investigating one famous context effect, one famous nudge and one set up more paradigmatic of the cue learning literature. We will demonstrate that in each case, the paradigmatically expected effect of the manipulation occurs in proportion to the probability with which the manipulated characteristic predicts the outcome of interest. Let us immediately state that this does not imply that previous accounts of these effects are “false”. Rather, it suggests that people are *also* able to adapt to achieve their goals and that this adaptation may determine what strategy is disclosed by a given task (cf.³⁴).

The first effect we will be focusing on is the attraction effect. The attraction effect exists when adding some dominated option (a decoy) to an option set increases the probability of a participant selecting the dominating option (the option that has a decoy) rather than a third option that dominates neither the decoy nor the option that has a decoy. This constitutes a violation of independence from irrelevant alternatives, which states that the choice between any pair of options should be unaffected by the addition or removal of other options to/from the option set³⁵. Although the attraction effect has been demonstrated many times (e.g.^{36–38}), failures to replicate the effect (e.g.³⁹) have led to studies qualifying the conditions under which it occurs (e.g.^{40–42}). Indeed, sometimes the effect of an option having a decoy is even reversed (“repulsion effect”,⁴³). It seems important to tune the differences between options (e.g.^{44,45}) such that they exist within some critical zone where the effect can be observed. There are (at least) two prominent psychological explanations of the effect: the decoy could provide individuals with a convenient argument for justifying their choice^{46–49} or it could arise from people evaluating the options through sequential comparisons of their attributes^{50,51}.

The second effect we will be focusing on is the default nudge (e.g.⁵²). A “nudge”²¹ is a manipulation of the context of a choice that does not restrict the freedom of choice of the individual. A famous example is to make some option the pre-selected default, which should make it more likely to be chosen over the alternatives (e.g.^{53,54}). However, in some settings this nudge has yielded null or even counter-predictive effects^{55–58}. A controversy surrounding a recent meta analysis of nudges

⁵⁹ has highlighted that the literature also suffers from severe publication bias^{60–62}. If some nudges are not robust it could thus be because they simply do not work in the first place.

In a given experiment investigating the attraction effect, it could turn out that the option that has a decoy tends to yield a superior, inferior, or on-par outcome compared to the alternatives. Similarly, in a default nudge study, it could turn out that the default option tends to be a “sensible” default⁶³ that yields the superior outcome. Participants may pick up and rely on such tendencies in order to make a choice, selecting for example the default option only in so far as it indeed tends to be the “sensible” choice. We propose that such a reliance can compete with — or even out-compete — the attraction effect and the effect of a default nudge as typically understood. If so, it suggests that the attraction effect and default nudge effect *per se* may only shine through under particular circumstances. Specifically, when the option that has a decoy/the default option tends to yield about as good

outcomes as the alternatives — a situation where the participant has reason to be indifferent. We think of these as empirical examples that motivate increased attention to how cue-outcome relationships may limit robustness.

Method

Participants performed 40 trials of a task previously used to study the attraction effect^{64,65}. Each participant was enrolled in one of three cue type conditions: the decoy, default, or rule condition. While we present these as three conditions within a single experiment, one could equally well think of them as three separate experiments. The decoy and default cue types are described in the Introduction. The rule cue type manipulated a physical stimulus characteristic, similarly to traditional cue learning literature (e.g.^{66,67}). This was intended as a comparison condition to see if the results from the two former cue types were comparable to this more traditional cue type, potentially illustrating that the former are cues “like any other”. The primary manipulation was how predictive the decoy/default/rule is (the probability of the cue indicating the superior option), which we randomised for each participant individually. We view each cue type condition as a replication of the effect of cue predictivity on choice behaviour.

Participants

Data was collected on Prolific, a popular online social science laboratory (see^{68,69} for overviews). Data for the decoy and default conditions was collected from September through October of 2021. Data for the rule condition was collected through March of 2024. We had a rolling recruitment with the preregistered goal of having at least 300 valid participants for each cue type. In total, we recruited 1010 participants. After applying our preregistered exclusion criterion of removing any participant who did not complete the experiment within one hour, the final sample contained 302 participants in the decoy condition, 306 in the default condition, and 301 in the rule condition. The average age of these participants was 29.3 years ($sd = 10.22$) and 47% were women.

Participation was open to all Prolific users who spoke fluent English, had participated in between 10 and 10 000 studies on Prolific, and had an approval rate of at least 95%. We did not specifically recruit from any vulnerable groups. We did not collect any demographic data on indicators of vulnerability due to ethical concerns — such data constitutes sensitive personal data, as per European and Swedish law, and must not be collected unless specifically motivated by the research question.

Participants collected a payoff (the area of their selected option minus its price) in experimental units in each round. After the experiment, the points were converted into GBP according to an exchange rate known to the subjects (1500 units to GBP 1). The average payment was GBP 2.85. The payment was calibrated to meet a fair standard of expected hourly pay.

Design

The task was based on Crosetto and Gaudeul⁶⁴ and Trueblood et al.⁶⁵. Participants were asked to choose one of three geometric shapes (each of which could be either a vertical rectangle, horizontal rectangle, vertical ellipse, horizontal ellipse, square, or circle) with different areas (e.g. 2 cm, 2.5 cm) presented on a grid. Each option came at some price, in experimental units. Upon choosing an option, the participant received a payoff equal to the area minus the price of the chosen option. That is, the immediately available information was sufficient to find the superior option (up to perceptual error); participants could disregard cue predictivity completely and still solve the task successfully. The participant also received full feedback⁷⁰: the areas of all shapes and their corresponding payoffs were presented on screen after each choice. At the end of the experiment, the experimental units were converted into pound sterling (GBP) at a predetermined exchange rate of 1500 units to GBP 1. Participants had to wait at least 5 seconds per trial before they could submit a choice, to invite some consideration. See Fig. 2 for task interface and feedback screen and Appendix A, Fig. S1, for task instructions.

We generalised the design of earlier decoy and default studies by generating option features randomly, from predetermined distributions, as follows: firstly, three payoffs were drawn from the same normal distribution and then ranked. Secondly, the price of each option was drawn from another normal distribution. Each option's area was determined by adding the payoff and the price. Thirdly, each option's geometric shape was randomised (but see the Rule condition below). These features exhaustively defined each option. In addition to this shared feature generation protocol, options were adjusted depending on cue type condition as described under the respective subheadings below. We manipulated cue predictivity between participants by randomising the probability of the cue indicating the superior option (“assigned predictivity”). In 28 of the rounds, the cue was generated according to the assigned predictivity (“cue treatment rounds”). In the remaining 12 rounds, the cue was unpredictable in expectation (uncorrelated with payoff) for all participants and identically generated for all participants within the same cue type condition. We used only these 12 test rounds to test our hypotheses, in order to respect measurement invariance^{71,72}. Treatment and test rounds were mixed and presented in random order. Option position (left, middle, or right) was randomised. In trials where the cue did not indicate the superior option, it indicated one of the other two options. That is: for each participant, there existed some individual probability that the option that has a decoy/default option/option of a particular shape (see conditions below) would also, on any given trial, be the option with the greatest potential payoff. So, on such trials the option that has a decoy, for example, would also turn out to be the option with the highest payoff (largest difference between area and price). On the rest of the trials, the option that has a decoy would not be the one with the highest payoff. There were no other differences between these two kinds of trials. If participants picked up on that having a decoy/being the default/being of a particular shape was a cue for being the superior option, they could use it to adapt their decision strategy. The probability, the relevant cue type, and, indeed, the existence of predictive cues were never stated to participants — they had to be discovered from experience. There were no visual indications of whether a round was a treatment or test round.

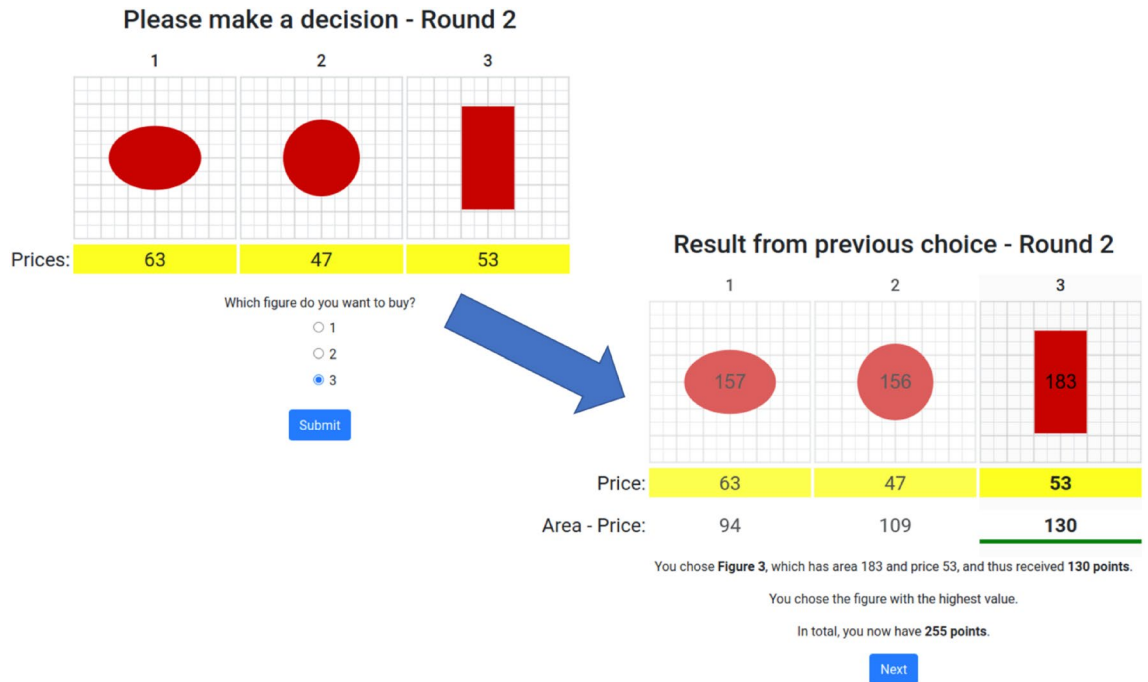


Fig. 2. The task as it appeared to participants. A participant in the default condition is about to choose option 3, a rectangle of area 183 and price 53. Once they have done so it is revealed that this yields 130 points. They are also informed that they chose the superior option, how many points the alternatives would have yielded, and their total score.

For any given trial, we indicated whether the cue did predict the superior option in that round using a dummy variable:

$$\text{cue predictive} = \begin{cases} 1 & \text{if superior option has a decoy/is default/is indicated by rule} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

We proceeded to calculate the share of such predictive rounds observed by the participant so far. We call that variable historical predictivity:

$$\text{historical predictivity}_r = \sum_{i=1}^{r-1} \frac{\text{cue predictive}_i}{r-1} \quad (2)$$

where r is rounds from 1 to 43, rounds 1 to 3 are practice rounds (see Procedure), and rounds 4 to 43 are the experiment rounds. This measure was the independent variable in our analyses. We thus did not investigate the effect of the *assigned* predictivity (cf. intention to treat⁷³) but of the empirical, historical predictivity that participants actually observed. We now describe how we adjusted the generated options for each cue type condition.

Decoy condition

In each of the 28 treatment rounds, option features were randomised (see Design) to generate three options. The generated options were adjusted as follows. With probability p equal to the assigned predictivity, the superior option was selected to have a decoy. If so, one of the remaining options (either the second or the third best) was chosen with 50% probability to be the superior option's decoy. Both the superior option and decoy option - the option that *has* a decoy and the option that *is* a decoy - were made to assume the same randomised shape. This facilitated ranking them by area (see⁷⁴⁻⁷⁶), making it slightly easier to identify that the decoy was dominated. The decoy option was then adjusted to be dominated by the superior option in both area and price as follows. Remember, the superior option would at this stage already, by definition, have a higher payoff than the decoy. We could thus calculate a payoff difference between the two. We adjusted the decoy by randomising a proportion $[0,1]$. That randomised proportion of the payoff difference was added to the superior option's price to generate the decoy's price. The rest of the payoff difference was deducted from the superior option's area to generate the decoy's area.

With probability $1 - p$, the second best option was chosen to have a decoy. If so, the worst option was the decoy and was adjusted as before. In each of the 12 test rounds, the generated options were adjusted as follows. Firstly, either the superior or the second-best option was determined to have a decoy with equal probability. That option was then assigned the worst option as its decoy. The decoy was adjusted as before. In the test rounds, an

option having a decoy was thus not indicative of whether it was the superior or second best option but it did indicate that it was not the worst option.

Default condition

In each of the 28 treatment rounds, option features were randomised (see Design) to generate three options. The generated options were adjusted as follows. With probability p equal to the assigned predictivity, the superior option of the three was pre-selected and participants thus only had to click the “submit” button to choose it. With probability $1 - p$, one of the two inferior options was pre-selected. In the 12 test rounds, one of the generated options was pre-selected by default with a uniform probability over the three options. In the test rounds, the default was thus not indicative of whether an option was superior, second best or third best.

Rule condition

For the 28 treatment rounds, one geometric shape per participant was randomised to be the “indicator shape” (e.g. square). Option features were randomised as described in Design to generate three options. With probability p the superior option was forced to assume the indicator shape and the shapes of the other two options were re-randomised to be any other shape. In $1 - p$ of cases, one of the two inferior options was instead forced to assume the indicator shape. The rule was thus that one of the shapes indicated the superior option with some probability p , and that shape would be present on every trial whether it was indicative or not. In the 12 test rounds, option features were generated as per Design and a randomly selected option was then forced to assume the indicator shape while the remaining two had their shapes re-randomised. In the test rounds, the indicator shape was thus not indicative of whether an option was superior, second best or third best.

Analyses, open science notes and ethics

We preregistered the experiment task, sample size, analyses and the hypotheses of positive associations between historical predictivity and following the decoy/default/rule in the test rounds.

To test these hypotheses, and as preregistered, we subsetted the data in the decoy, default, and rule condition, respectively, to obtain only the 12 test rounds per condition. Separately for each condition, we then regressed the binary variable of whether the option with a decoy/default option/option indicated by the rule was chosen on the historical predictivity. For each regression, we predicted a positive coefficient of the historical predictivity measure.

According to the Swedish act concerning the Ethical Review of Research involving humans (2003:460), blanket ethics approval applied to this study. The study was conducted in accordance with the Declaration of Helsinki.

Procedure

Participants gave informed consent (Fig. S2), received identical instructions in all conditions, and were able to practice the task for three rounds. The practice rounds were generated according to the participants’ respective treatment levels: their cue type and their assigned predictivity. Participants then needed to pass a comprehension quiz with 4 questions (Fig. S3). If they failed it more than 10 times, their participation was terminated. If they passed, participants proceeded through the 40 experiment rounds. After the experiment, participants were compensated for their time.

Results

We find evidence of positive effects of historical predictivity in our preregistered regressions for all cue types, see Table 1. An increase of historical predictivity by 10 percentage points results in an estimated increase in the probability of choosing the option that has a decoy, the default, and the option indicated by the rule by 1.29%, 1.58%, and 2.21%, respectively.

We can graph the estimated linear relationship between the mean probability of choosing the option that has a decoy/is the default/is indicated by the rule and the historical predictivity, see Fig. 3. What the relationships should look like in the presence or absence of an adaptation to cue predictivity differs slightly between conditions. We walk through those predictions in Appendix B and state the take-home message here: these graphs show that when the cue has historically been predictive, participants choose the option that has a decoy/is the default/

	Dependent variable		
	Option with decoy was chosen	Default option was chosen	Option indicated by rule was chosen
	(1)	(2)	(3)
Constant	0.476 (0.021)	0.283 (0.020)	0.265 (0.020)
Historical predictivity	0.129 (0.037)	0.158 (0.039)	0.221 (0.044)

Table 1. Regression parameter values for each dependent variable. Standard errors in parentheses, robust and clustered on individual level. $p < 0.001$ for all estimates. In Appendix C, Table S1, we show that these results are robust to including participant level random effects instead of the clustered standard errors.

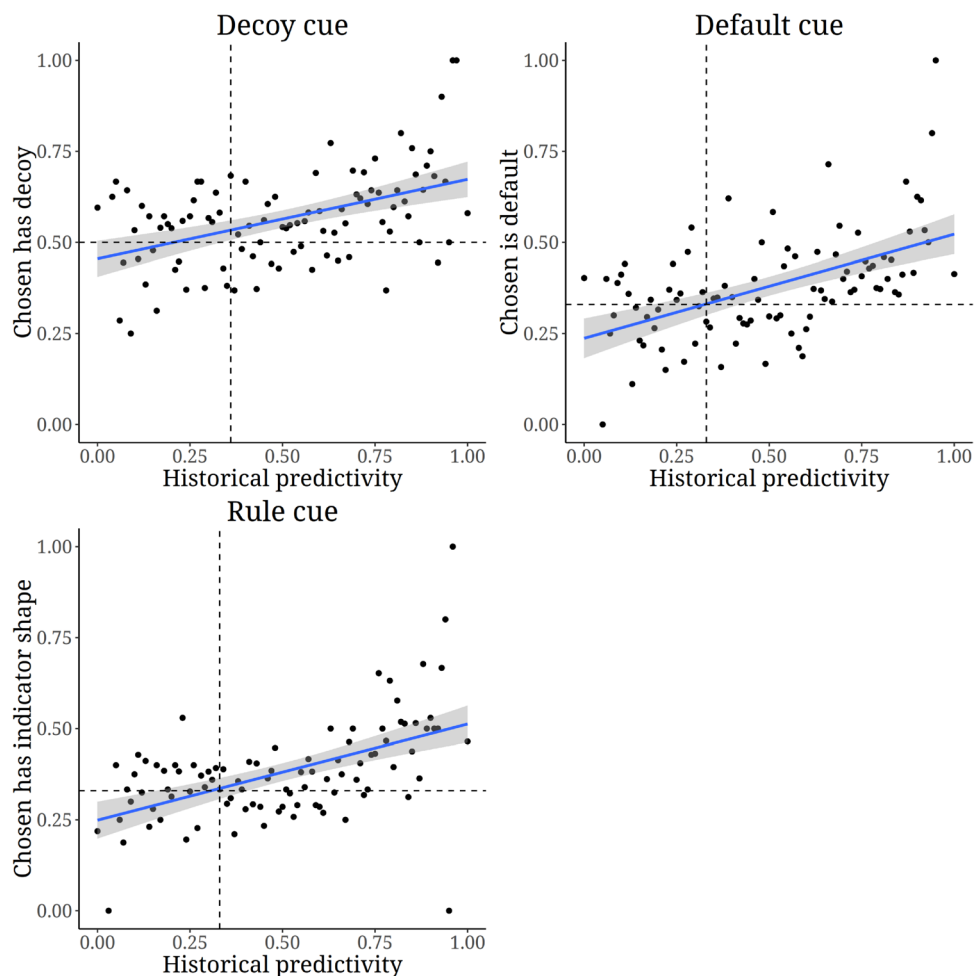


Fig. 3. Regressions of historical predictivity on choosing the option that has a decoy/is the default/is of the indicator shape. The latter are binary variables, making individual observations difficult to visualise in a scatterplot since they will all sit at 0 or 1. We therefore group all observations by historical predictivity rounded to two decimal places. Each marker is thus a local average for a small subset of observations at each level of historical predictivity. Intersections of dashed lines indicate at what level of cue predictivity choice probabilities should be at chance if participants respond to the cues (see Appendix B, Fig. S4). The regression lines should thus intersect these intersections.

is indicated by the rule above chance frequency. When the cue has not been predictive, there is a tendency towards a slight attraction effect but choice in the default condition is on average practically random. When the cue has been predictive of which option *is not* superior, the superior option is on average chosen below chance frequency. Indeed, if we follow the convention of evaluating these effects based on group averages (contrast this to⁷⁷) we can find cross sections where the effects appear, disappear, and (less clearly) are reversed.

The estimated cue weight (coefficient of historical predictivity, Table 1) appears to be greatest for the rule condition, second greatest for the default condition, and smallest for the decoy condition. A tentative speculation is that this reflects how complex the cues are: the rule involves a property of the stimulus, the default involves an association between the response mode and the stimulus, while the decoy involves a relationship between two stimuli. The reliance on cues could be moderated by how easy they are to pick up.

Discussion

Our most fundamental conclusion is that the prevalence of the effects investigated here scaled with how strongly the cues provided by characteristics of the experimental manipulations had predicted the superior option. Less clearly, when the manipulations helped participants predict which option was *not* superior, the directions of the effects seemed to reverse. The effects we focused on here were thus not robust to cue-outcome relationships. As an anonymous reviewer pointed out, a probabilistic functionalist¹¹ or (social psychological) judgement correction researcher (e.g.^{23,26,24}) might find these results highly expected; established principles applied to a novel setting. We agree! Despite this, these established principles seem (to us) to not have permeated into the broader social scientific and meta-scientific discourse. It does not seem recognised that they can limit robustness in the way they did here. We tentatively suggest that this perspective may be important for such discussions.

One way to interpret the results is that cue predictivity is a potential confound of the substantive effects. If so, the question becomes how to avoid the confounding. We would perhaps prefer a different interpretation: all effects are contextual in the sense that they are moderated by other variables than the main variable of interest (cf.⁶). There is no pure or “true” effect of X on Y , only a conditional effect $Y = f(X_1|X_2, X_3, \dots, X_n)$ (cf.⁷⁸). A robust effect would be one where these moderators X_2, X_3, \dots, X_n exert only a minor influence on the final result. Here we have demonstrated a moderator that has appreciable influence, but the solution is not to try to do to away with it but to embrace it. Any investigation of, say, the attraction effect will reveal that effect *conditional on a particular level of cue predictivity*. That is absolutely fine, as long as we interpret the results in that manner and carry that interpretation with us as we draw theoretical and applied conclusions. If we have some context in mind that we want our results to hold in, we should set the cue predictivities in our study such that they are representative of that setting (cf. “representative design”^{79,80}).

Given that view, what do these results mean for the literatures on the attraction effect and default nudge? Firstly, the sometimes failed attempts at replication of the attraction effect (see Introduction) indicate that whatever cognitive mechanism explains it is not ubiquitous. While we encourage independent replications, if we take the present results as read they further circumscribe such a mechanism: a theory that explains the attraction effect (see e.g. General discussion in⁶⁵) will not be a theory of “how people make choices” but rather of “how some people make choices, at least when there are no highly predictive cues”. To be very clear: this does not make such theories less interesting, exciting, or worthwhile. However, we agree with Regenwetter, Robinson, and Wang⁸¹ that the theoretical scope should be made explicit. Secondly, effects of defaults have been proclaimed to be due to people using a “default heuristic” (e.g.⁸²). Labelling something a “heuristic” is, on its own, a redescription rather than an explanation (cf.⁸³) but in this case the suggestion seems to be that there exists some more or less hard wired algorithm that people are wont to use (see⁸⁴). We suggest that it might be more fruitful to proclaim the default a “cue” and (re)describe default effects as the cue having a “high cue weight” because this situates the phenomenon within a principled theoretical framework. Such a framework might then open up for new explanations. Specifically that, far from being hard wired, default effects are emergent properties of the individual attuning to their environment. Indeed, here we saw no default effect at all on average when historical cue predictivity was about chance level (Fig. 3). It might also open up new avenues of empirical inquiry. For example, Brunswikian psychology has a long history of lens model analysis (e.g.^{85,86}) where cue predictivity, cue weight, and achievement are estimated empirically for a given setting. Such an exercise sheds light on both the normative (cue predictivity) and empirical (cue weight) solution, as well as raw performance (achievement).

One could interpret these results in light of the literature on probability learning. A now classic review by Peterson and Beach⁸⁷ presented people as “intuitive statisticians” who are able to learn statistics of observed samples of outcomes (see e.g.^{88,289–91} for contemporary work). When making discrete decisions based on such experience, participants are often found to “probability match”^{92–95} - they select the superior option proportionally to how often it yielded the superior outcome. This deviates from the normatively optimal strategy of “maximisation” - to always select the option most likely to yield the superior outcome - but has been argued to be superior in an ecology where seemingly stochastic payoffs may in fact reflect sequential dependencies⁹⁶. One could interpret Fig. 3 as showing a tendency towards probability matching here too. That being said, such a discussion is not central to our present claim: it is the *function* rather than the *mechanism* (see Tinbergen’s levels of analysis,⁹⁷) that motivates our limiting condition for robustness.

What are the limiting conditions of this limiting condition? First and foremost, cue-outcome relationships must exist and be strong enough to be picked up. Mere repetition is not sufficient. People can learn cue weights or make one-shot decisions using inferred ones, and it does not seem to matter much if the relevant cues are stated explicitly by an experimenter or if the individual has to extract them without guidance¹³, as in this study. Secondly, relying on the cue must improve on other grounds for choice (cf. “cue family hierarchy”^{98,99}). In the present experiment, participants were provided with all the information necessary to, at least in principle, always identify the superior option. Had we made the task very easy, by greatly increasing the differences in area and price between options, a probabilistic functionalist would expect participants to ignore the cue and instead base their decision purely on the option attributes. Thirdly, we agree with flexible correction model theorists (e.g.^{26,24}) that one must be motivated to achieve the outcome a given cue predicts for it to affect choice behaviour. However, we emphasise that those goals can be highly idiosyncratic (e.g.¹⁰⁰), although incentivisation can reduce that heterogeneity (e.g.¹⁰¹).

Whether the limiting condition we have argued for here is an important one hinges on there existing cue-outcome relationships in the settings where we “want” phenomena to be robust (cf.¹). We tentatively believe such settings to exist. Nudges provide a case in point (see also¹⁰²). The nudge will, again, involve the creation of cues. It will be applied with the aim of moving individuals towards making a particular choice which begets a particular outcome. This constitutes a cue-outcome relationship. When the choice architect tries to nudge the individual towards an outcome that is not aligned with the latter’s goals, the individual may eventually recalibrate their cue weights to disregard the nudge. Our limiting condition could thereby come into play here; it could be that nudges have more robust effects when they nudge an individual *towards* her self-perceived goals (“motivational alignment”¹⁰³). On the flip side, this limitation could also be a boon. Authorities and so-called “nudge units” could potentially exploit people’s capacity to adapt to cue-outcome relationships to “boost”¹⁰⁴ individuals in at least two ways. Firstly, they could ensure motivational alignment between nudgers and nudgees and then construct highly predictive cues for those aligned goals. Alternatively, they could ensure that there exist highly predictive cues for a diverse range of outcomes. That way, one might not need to infer the individual’s self-perceived goals. Instead, one might trust individuals to pick up on the relevant cues and rely on them to when making their choices. In Brunswikian terms, this would allow individuals to obtain high achievement (Fig. 1) to whatever outcome they desire. While speculative, we believe this to be an exciting prospect for future research to address.

Data availability

The decoy and default conditions were preregistered at <https://doi.org/10.17605/OSF.IO/DS4XP>. Data for the rule condition was collected later and preregistered at <https://doi.org/10.17605/OSF.IO/852ZV>. A preregistered pilot experiment is available in Mandl (2022). All data, materials and analysis code are available at <https://doi.org/10.17605/OSF.IO/J362H>.

Received: 9 May 2025; Accepted: 10 December 2025

Published online: 22 December 2025

References

- Nancy, Cartwright. & Sophia, Efstathiou. "Hunting Causes and Using Them: Is There No Bridge from Here to There?" In: *International Studies in the Philosophy of Science* 25.3 (Sept. 2011), 223–241. issn: 1469-9281. <https://doi.org/10.1080/02698595.2011.605245>.
- Mattias, Forsgren, Peter, Juslin, & Ronald, van den Berg. "Further perceptions of probability: Accurate, stepwise updating is contingent on prior information about the task and the response mode". In: *Psychonomic Bulletin & Review* (Nov. 2024). issn: 1531-5320. <https://doi.org/10.3758/s13423-024-02604-2>.
- Christophe, Gernigon et al. "How the Complexity of Psychological Processes Reframes the Issue of Reproducibility in Psychological Science". In: *Perspectives on Psychological Science* 19.6 (Aug. 2023), 952–977. issn: 1745-6924. <https://doi.org/10.1177/17456916231187324>.
- Brian, A. Nosek et al. "Replicability, Robustness, and Reproducibility in Psychological Science". In: *Annual Review of Psychology* 73.1 (Jan. 2022), 719–748. issn: 1545-2085. <https://doi.org/10.1146/annurev-psych-020821-114157>.
- Open Science Collaboration. "Estimating the reproducibility of psychological science". In: *Science* 349.6251 (Aug. 2015). issn: 1095-9203. <https://doi.org/10.1126/science.aac4716>.
- Andrew, Gelman. "The Connection Between Varying Treatment Effects and the Crisis of Unreplicable Research: A Bayesian Perspective". In: *Journal of Management* 41.2 (Mar. 2014), 632–643. issn: 1557-1211. <https://doi.org/10.1177/0149206314525208>.
- Antonia, Krefeld-Schwalb, Eli, Rosen Sugerman, & Eric, J. Johnson. "Exposing omitted moderators: Explaining why effect sizes differ in the social sciences". In: *Proceedings of the National Academy of Sciences* 121.12 (Mar. 2024). issn: 1091-6490. <https://doi.org/10.1073/pnas.2306281121>.
- Christopher, Tosh et al. "The Piranha Problem: Large Effects Swimming in a Small Pond". In: *Notices of the American Mathematical Society* 72.01 (Jan. 2025), p.1. issn: 1088-9477. <https://doi.org/10.1090/noti3044>.
- Egon, Brunswik. *The conceptual framework of psychology*. eng. International encyclopaedia of unified science, 1:10. Chicago: Univ. of Chicago Press, (1952).
- Egon, Brunswik. *Perception and the representative design of psychological experiments*. 2nd. University of California Press, (1956).
- Egon, Brunswik. "In defense of probabilistic functionalism: a reply." In: *Psychological Review* 62.3 (1955), 236–242. issn: 0033-295X. <https://doi.org/10.1037/h0040198>.
- Carolyn, J. Hirsch, Kenneth, R. Hammond, & Jack, L. Hirsch. "Some methodological considerations in multiple-cue probability studies." In: *Psychological Review* 71.1 (Jan. 1964), 42–60. <https://doi.org/10.1037/h0041729>.
- Natalia, Karelaia, & Robin, M. Hogarth. "Determinants of linear judgment: A meta-analysis of lens model studies." In: *Psychological Bulletin* 134.3 (2008), 404–426. <https://doi.org/10.1037/0033-2909.134.3.404>.
- Peter, Juslin et al. "Cue abstraction and exemplar memory in categorization." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29 (5 2003), 924–941. issn: 1939-1285. <https://doi.org/10.1037/0278-7393.29.5.924>.
- R. Shayna Rosenbaum, Alice S.N. Kim, & Stevenson Baker. "2.06 - Episodic and Semantic Memory". In: *Learning and Memory: A Comprehensive Reference (Second Edition)*. Ed. by John H. Byrne. Second Edition. Oxford: Academic Press, 2017, 87–118. isbn: 978-0-12-805291-4. <https://doi.org/10.1016/B978-0-12-809324-5.21037-7>. <https://www.sciencedirect.com/science/article/pii/B9780128093245210377>.
- Tulving, Endel *Episodic and semantic memory*. In *Organization of memory* (eds Tulving, Endel & Donaldson, Wayne) 381–403 (Academic Press, New York, 1972).
- Amos, Tversky. "Elimination by aspects: A theory of choice." In: *Psychological Review* 79.4 (July 1972), 281–299. issn: 0033-295X. <https://doi.org/10.1037/h0032955>.
- Pedro, Bordalo, Nicola, Gennaioli. & Andrei, Shleifer. "Salience". In: *Annual Review of Economics* 14.1 (Aug. 2022), 521–544. issn: 1941-1391. <https://doi.org/10.1146/annurev-economics-051520-011616>.
- Maya, Bar-Hillel. "Position Effects in Choice From Simultaneous Displays: A Conundrum Solved". In: *Perspectives on Psychological Science* 10.4 (July 2015), 419–433. issn: 1745-6924. <https://doi.org/10.1177/1745691615588092>.
- Peter, Juslin, Henrik, Olsson, & Anna-Carin, Olsson. "Exemplar effects in categorization and multiple-cue judgment." In: *Journal of Experimental Psychology: General* 132.1 (2003), 133–156. issn: 0096-3445. <https://doi.org/10.1037/0096-3445.132.1.133>.
- Richard, H. Thaler, & Cass, Sunstein. "Nudge: Improving decisions about health, wealth and happiness". In: *Amsterdam Law Forum; HeinOnline: Online. HeinOnline. 89, (2008)*.
- G.Elliott, Wimmer, & Daphna, Shohamy. "Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions". In: *Science* 338.6104 (Oct. 2012), 270–273. issn: 1095-9203. <https://doi.org/10.1126/science.1223252>.
- Norbert, Schwarz, & Herbert, Bless. "Mental construal processes: The inclusion/exclusion model". In: *Assimilation and Contrast in Social Psychology*. Ed. by Diederik A. Stapel and Jerry M. Suls. New York, NY: Psychology Press, 2007, 119–141. isbn: 978-1-84169-449-8.
- Duane, T. Wegener, & Richard, E. Petty. "The Flexible Correction Model: The Role of Naive Theories of Bias in Bias Correction". In: *Advances in Experimental Social Psychology* Volume 29. Elsevier, 141–208. [https://doi.org/10.1016/s0065-2601\(08\)60017-9](https://doi.org/10.1016/s0065-2601(08)60017-9) (1997).
- Steven, Sloman. *Causal Models: How People Think about the World and Its Alternatives*. Oxford University Press New York, Aug. . isbn: 9780199870950. (2005) <https://doi.org/10.1093/acprof:oso/9780195183115.001.0001>.
- Duane, T. Wegener, & Richard, E. Petty. "Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias." In: *Journal of Personality and Social Psychology* 68.1 (1995), 36–51. issn: 0022-3514. <https://doi.org/10.1037/0022-3514.68.1.36>.
- Gerd, Gigerenzer, Ulrich, Hoffrage, & Heinz, Kleinbölting. "Probabilistic mental models: A Brunswikian theory of confidence." In: *Psychological Review* 98.4 (1991), 506–528. issn: 0033-295X. <https://doi.org/10.1037/0033-295X.98.4.506>.
- Robin, M. Hogarth, & Natalia Karelaia. " "Take-the-Best" and Other Simple Strategies: Why and When they Work "Well" with Binary Cues". In: *Theory and Decision* 61.3 (Sept. 2006), 205–249. issn: 1573-7187. <https://doi.org/10.1007/s11238-006-9000-8>.
- Vladimir, Chituc, Crockett, M.J. & Brian, J. Scholl. "How to show that a cruel prank is worse than a war crime: Shifting scales and missing benchmarks in the study of moral judgment". In: *Cognition* 266 (Jan. 2026), p.106315. issn: 0010-0277. <https://doi.org/10.1016/j.cognition.2025.106315>.

30. Leonard, Berkowitz, & Edward, Donnerstein. "External validity is more than skin deep: Some answers to criticisms of laboratory experiments." In: *American Psychologist* 37.3 (Mar. 1982), 245–257. issn: 0003-066X. <https://doi.org/10.1037/0003-066X.37.3.245>.
31. Nancy, Cartwright. "Knowing what we are talking about: why evidence doesn't always travel". In: *Evidence and Policy* 9.1 (Jan. 2013), 97–112. issn: 1744-2656. <https://doi.org/10.1332/174426413x662581>.
32. Gerd, Gigerenzer. "External Validity of Laboratory Experiments: The Frequency-Validity Relationship". In: *The American Journal of Psychology* 97.2 (1984), 185. issn: 0002-9556. <https://doi.org/10.2307/1422594>.
33. John, A. List. "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions". In: *Journal of Political Economy* 114.1 (2006), 1–37. <https://doi.org/10.1086/498587>.
34. Julian, N. Marewski, & Lael, J. Schooler. "Cognitive niches: An ecological model of strategy selection." In: *Psychological Review* 118.3 (2011), 393–437. issn: 0033-295X. <https://doi.org/10.1037/a0024143>.
35. Gerard, Debreu. *Individual choice behavior: A theoretical analysis*. (1960).
36. Geoffrey, Castillo. "The Attraction Effect and Its Explanations". In: *Games and Economic Behavior* 119 (Jan. 2020), 123–147. issn: 08998256. <https://doi.org/10.1016/j.geb.2019.10.012>. <https://linkinghub.elsevier.com/retrieve/pii/S0899825619301654> (visited on 08/24/2020).
37. George, D. Farmer et al. "The Effect of Expected Value on Attraction Effect Preference Reversals". In: *Journal of Behavioral Decision Making* 30.4 (Dec. 2016), 785–793. issn: 1099-0771. <https://doi.org/10.1002/bdm.2001>.
38. Holger, Müller, Victor, Schliwa, & Sebastian Lehmann. "Prize Decoys at Work — New Experimental Evidence for Asymmetric Dominance Effects in Choices on Prizes in Competitions". In: *International Journal of Research in Marketing* 31.4 (2014), 457–460. issn: 01678116. <https://doi.org/10.1016/j.ijresmar.2014.09.003>. (visited on 03/27/2020).
39. Sybil, Yang, & Michael, Lynn. "More Evidence Challenging the Robustness and Usefulness of the Attraction Effect". en. In: *Journal of Marketing Research* 51.4 (Aug. 2014). Publisher: SAGE Publications Inc, 508–513. issn: 0022-2437. <https://doi.org/10.1509/jmr.14.0020>. (visited on 10/18/2021).
40. Paolo, Crosetto, & Alexia, Gaudeul. "Fast Then Slow: Choice Revisions Drive a Decline in the Attraction Effect". In: *Management Science* 70.6 (June 2024), 3711–3733. issn: 1526-5501. <https://doi.org/10.1287/mnsc.2023.4874>.
41. Shane, Frederick, Leonard, Lee, & Ernest, Baskin. "The Limits of Attraction". en. In: *Journal of Marketing Research* 51.4 (Aug. 2014). Publisher: SAGE Publications Inc, 487–507. issn: 0022-2437. <https://doi.org/10.1509/jmr.12.0061>. (visited on 10/18/2021).
42. Marcel, Lichters et al. "What Really Matters in Attraction Effect Research: When Choices Have Economic Consequences". In: *Marketing Letters* 28.1 (Mar. 2017), 127–138. issn: 0923-0645, 1573-059X. <https://doi.org/10.1007/s11002-015-9394-6>. (visited on 08/24/2020).
43. Pronobesh, Banerjee et al. "Factors that promote the repulsion effect in preferential choice". In: *Judgment and Decision Making* 19 (2024). issn: 1930-2975. <https://doi.org/10.1017/jdm.2023.46>.
44. Maurits, C Kaptein, Robin, Van Emden, & Davide, Iannuzzi. "Tracking the Decoy: Maximizing the Decoy Effect through Sequential Experimentation". In: *Palgrave Communications* 2.1 (Dec. 2016), p.16082. issn: 2055-1045. <https://doi.org/10.1057/palcomms.2016.82>. <http://www.nature.com/articles/palcomms201682> (visited on 08/23/2020).
45. Pravesh, Kumar Padamwar, Jagrook, Dawra, & Vinay, Kumar Kalakbandi. "The Impact of Range Extension on the Attraction Effect". In: *Journal of Business Research* (Dec. 2019), S0148296319307830. issn: 01482963. <https://doi.org/10.1016/j.jbusres.2019.12.017>. <https://linkinghub.elsevier.com/retrieve/pii/S0148296319307830> (visited on 03/27/2020).
46. Franz, Dietrich, & Christian, List. "Reason-based choice and context-dependence: An explanatory Framework". en. In: *Economics and Philosophy* 32.2 (July 2016), 175–229. issn: 0266-2671, 1474-0028. <https://doi.org/10.1017/S0266267115000474>. https://www.cambridge.org/core/product/identifier/S0266267115000474/type/journal_article (visited on 03/30/2020).
47. Yolanda Gomez et al. "The Attraction Effect in Mid-Involvement Categories: An Experimental Economics Approach". In: *Journal of Business Research* 69.11 (Nov. 2016), pp.5082–5088. issn: 01482963. <https://doi.org/10.1016/j.jbusres.2016.04.084>. <https://linkinghub.elsevier.com/retrieve/pii/S0148296316302478> (visited on 03/27/2020).
48. Eldar, Shafir, Itamar, Simonson, & Amos Tversky. "Reason-Based Choice". In: *Cognition* 49.1-2 (Oct.–Nov. 1993), 11–36.
49. Itamar, Simonson. "Choice Based on Reasons: The Case of Attraction and Compromise Effects". In: *Journal of Consumer Research* 16.2 (Sept. 1989), 158. <https://doi.org/10.1086/209205>.
50. Robert, M. Roe, Jermon, R. Busemeyer, & James, T. Townsend. "Multialternative decision field theory: A dynamic connectionist model of decision making." In: *Psychological Review* 108.2 (2001), 370–392. <https://doi.org/10.1037/0033-295x.108.2.370>.
51. Jennifer, S. Trueblood, Scott, D. Brown, & Andrew, Heathcote. "The multiattribute linear ballistic accumulator model of context effects in multialternative choice." In: *Psychological Review* 121.2 (Apr. 2014), 179–205. <https://doi.org/10.1037/a0036137>.
52. David, Tannenbaum, Craig, R. Fox, & Todd Rogers. "On the Misplaced Politics of Behavioural Policy Interventions". In: *Nature Human Behaviour* 1.7 (July 2017), p.0130. issn: 2397-3374. <https://doi.org/10.1038/s41562-017-0130>. (visited on 04/14/2020).
53. Cronqvist, Henrik, Thaler, Richard H. & Frank, Yu. When Nudges Are Forever: Inertia in the Swedish Premium Pension Plan. *AEA Papers and Proceedings* 108, 153–158. <https://doi.org/10.1257/pandp.20181096>.
54. Barnabas, Szaszi et al. "A Systematic Scoping Review of the Choice Architecture Movement: Toward Understanding When and Why Nudges Work". en. In: *Journal of Behavioral Decision Making* 31.3 (2018). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/bdm.2035>, 355–366. issn: 1099-0771. <https://doi.org/10.1002/bdm.2035>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/bdm.2035> (visited on 10/18/2021).
55. Steffen, Altmann et al. "Defaults and Donations: Evidence from a Field Experiment". In: *The Review of Economics and Statistics* 101.5 (Dec. 2019), 808–826. issn: 0034-6535. https://doi.org/10.1162/rest_a_00774 (visited on 10/19/2021).
56. John, Beshears et al. The Limitations of Defaults. en. 2010. <https://www.nber.org/programs-projects/projects-and-centers/retireent-and-disability-research-center/papers/rrc-nb10-02> (visited on 10/19/2021).
57. Erin, Bronchetti et al. "When a Nudge Isn't Enough: Defaults and Saving Among Low-Income Tax Filers". In: *National Tax Journal* 66 (Mar. 2011). <https://doi.org/10.17310/ntj.2013.3.04>.
58. Jon M. Jachimowicz et al. "When and Why Defaults Influence Decisions: A Meta-Analysis of Default Effects". In: *Behavioural Public Policy* 3.2 (Nov. 2019), pp.159–186. issn: 2398-063X, 2398-0648. <https://doi.org/10.1017/bpp.2018.43>. <https://www.cambridge.org/core/journals/behavioural-public-policy/article/when-and-why-defaults-influence-decisions-a-metaanalysis-of-default-effects/67AF6972CFB52698A60B6BD94B70C2C0> (visited on 04/02/2020).
59. S. Mertens, M. Herberz, U.J.J. Hahnel, & T. Brosch. The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proc. Natl. Acad. Sci. U.S.A.* 119 (1) e2107346118, <https://doi.org/10.1073/pnas.2107346118> (2022).
60. Jonathan, Z Bakdash, & Laura, R Marusich. "Left-truncated effects and overestimated meta-analytic means". In: *Proceedings of the National Academy of Sciences* 119.31 (July 2022). issn: 1091-6490. <https://doi.org/10.1073/pnas.2203616119>.
61. Stephanie, Mertens et al. "Reply to Maier et al., Szaszi et al., and Bakdash and Marusich: The present and future of choice architecture research". In: *Proceedings of the National Academy of Sciences* 119.31 (July 2022). issn: 1091-6490. <https://doi.org/10.1073/pnas.2202928119>.
62. Barnabas, Szaszi et al. "No reason to expect large and consistent effects of nudge interventions". In: *Proceedings of the National Academy of Sciences* 119.31 (July 2022). issn: 1091-6490. <https://doi.org/10.1073/pnas.2200732119>.
63. Richard, H. Thaler, Cass, R. Sunstein, & John P. Balz. "Choice architecture". In: *The Behavioral Foundations of Public Policy*. Ed. by Eldar Shafir. Princeton, NJ: Princeton University Press, 428–439. isbn: 9780691137568, 2013.
64. Paolo, Crosetto, & Alexia, Gaudeul. "A Monetary Measure of the Strength and Robustness of the Attraction Effect". In: *Economics Letters* 149 (Dec. 2016), 38–43. issn: 01651765. <https://doi.org/10.1016/j.econlet.2016.09.031>. <https://linkinghub.elsevier.com/retrieve/pii/S0165176516303901> (visited on 08/24/2020).

65. Jennifer, S. Trueblood et al. “Not Just for Consumers: Context Effects Are Fundamental to Decision Making”. In: *Psychological Science* 24.6 (June 1, 2013), 901–908. issn: 0956-7976. <https://doi.org/10.1177/0956797612464241>. (visited on 10/20/2021).
66. Peter, Juslin, Linnea, Karlsson, & Henrik, Olsson. “Information integration in multiple cue judgment: A division of labor hypothesis”. In: *Cognition* 106.1 (Jan. 2008), 259–298. issn: 0010-0277. <https://doi.org/10.1016/j.cognition.2007.02.003>.
67. Dries, Trippas, Thorsten, Pachur. “Nothing compares: Unraveling learning task effects in judgment and categorization”. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 45.12 (Dec. 2019), 2239–2266. issn: 0278-7393. <https://doi.org/10.1037/xlm0000696>.
68. Stefan, Palan, & Christian, Schitter. “Prolific.ac—A subject pool for online experiments”. In: *Journal of Behavioral and Experimental Finance* 17 (Mar. 2018), 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>.
69. Peer, Eyal et al. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006> (2017).
70. Adrian, R. Camilleri & Ben, R. Newell. “When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms”. In: *Psychonomic Bulletin & Review* 18 (2 Apr. 2011), 377–384. issn: 1069-9384. <https://doi.org/10.3758/s13423-010-0040-2>.
71. Quentin, André, & Bart de, Langhe. “No evidence for loss aversion disappearance and reversal in Walasek and Stewart (2015)”. In: *Journal of Experimental Psychology: General* 150.12 (Dec. 2021), 2659–2665. issn: 0096-3445. <https://doi.org/10.1037/xge001052>.
72. William, Meredith. “Measurement invariance, factor analysis and factorial invariance”. In: *Psychometrika* 58 (4 Dec. 1993), 525–543. issn: 0033-3123. <https://doi.org/10.1007/BF02294825>. <http://link.springer.com/10.1007/BF02294825>.
73. Joshua, D. Angrist & Jörn-Steffen, Pischke. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press, Dec. 2009. isbn: 9781400829828. <https://doi.org/10.1515/9781400829828>.
74. Paulo, Natenzon. “Random choice and learning”. In: *Journal of Political Economy* 127.1 (2019), 419–457. issn: 1537534X. <https://doi.org/10.1086/700762>.
75. Veronica, Pisu et al. “Biases in the perceived area of different shapes: A comprehensive account and model”. In: *Journal of Experimental Psychology: Human Perception and Performance* 51.9 (Sept. 2025), 1167–1177. issn: 0096-1523. <https://doi.org/10.1037/xhp0001322>.
76. John, P. Smith. “The effects of figurai shape on the perception of area”. In: *Perception & Psychophysics* 5.1 (Jan. 1969), 49–52. <https://doi.org/10.3758/bf03210480>.
77. Håkan, Fischer, Mats, E. Nilsson, & Natalie, C. Ebner. “Why the Single-N Design Should Be the Default in Affective Neuroscience”. In: *Affective Science* 5.1 (Mar. 2023), 62–66. issn: 2662-205X. <https://doi.org/10.1007/s42761-023-00182-5>.
78. Armin, Falk, & James, J. Heckman. “Lab Experiments Are a Major Source of Knowledge in the Social Sciences”. In: *Science* 326.5952 (Oct. 2009), 535–538. issn: 1095-9203. <https://doi.org/10.1126/science.1168244>.
79. Mandeep, K. Dhami, Ralph, Hertwig, & Ulrich, Hoffrage. “The Role of Representative Design in an Ecological Approach to Cognition”. In: *Psychological Bulletin* 130 (6 2004), 959–988. issn: 1939-1455. <https://doi.org/10.1037/0033-2909.130.6.959>.
80. Gijs A. Holleman et al. “Representative design: A realistic alternative to (systematic) integrative design”. In: *Behavioral and Brain Sciences* 47 (2024). issn: 1469-1825. <https://doi.org/10.1017/s0140525x23002200>.
81. Michel, Regenwetter, Maria, M. Robinson, & Cihang, Wang. “Four Internal Inconsistencies in Tversky and Kahneman’s (1992) Cumulative Prospect Theory Article: A Case Study in Ambiguous Theoretical Scope and Ambiguous Parsimony”. In: *Advances in Methods and Practices in Psychological Science* 5.1 (Jan. 2022). issn: 2515-2467. <https://doi.org/10.1177/25152459221074653>.
82. Gerd, Gigerenzer, & Wolfgang, Gaissmaier. “Heuristic Decision Making”. In: *Annual Review of Psychology* 62.1 (Jan. 2011), 451–482. issn: 1545-2085. <https://doi.org/10.1146/annurev-psych-120709-145346>.
83. Gerd, Gigerenzer. “Surrogates for Theories”. In: *Theory & Psychology* 8.2 (Apr. 1998), 195–204. issn: 1461-7447. <https://doi.org/10.1177/0959354398082006>.
84. Gerd, Gigerenzer. “Why Heuristics Work”. In: *Perspectives on Psychological Science* 3.1 (2008), 20–29. issn: 17456916, 17456924. <http://www.jstor.org/stable/40212224> (visited on 10/13/2025).
85. N. John, Castellan. “Comments on the “Lens Model” Equation and Analysis of Multiple-Cue Judgment Tasks”. In: *Psychometrika* 38.1 (Mar. 1973), 87–100. issn: 1860-0980. <https://doi.org/10.1007/BF02291177>.
86. Reid, Hastie, & Kenneth, R. Hammond. “Rational analysis and the Lens model”. In: *Behavioral and Brain Sciences* 14.3 (Sept. 1991), 498–498. issn: 1469-1825. <https://doi.org/10.1017/S0140525X00070953>.
87. Cameron, R. Peterson, & Lee, Roy Beach. “Man as an intuitive statistician”. In: *Psychological Bulletin* 68.1 (July 1967), 29–46. issn: 0033-2909. <https://doi.org/10.1037/h0024722>.
88. Mattias, Forsgren, Peter, Juslin, & Ronald, van den Berg. “Further perceptions of probability: In defence of associative models.” In: *Psychological Review* 130.5 (Oct. 2023), 1383–1400. issn: 0033-295X. <https://doi.org/10.1037/rev0000410>.
89. Joseph, T. McGuire et al. “Functionally Dissociable Influences on Learning Rate in a Dynamic Environment”. In: *Neuron* 84.4 (Nov. 2014), 870–881. issn: 0896-6273. <https://doi.org/10.1016/j.neuron.2014.10.013>.
90. Matthew, R. Nassar et al. “An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment”. In: *The Journal of Neuroscience* 30.37 (Sept. 2010), 12366–12378. issn: 1529-2401. <https://doi.org/10.1523/jneurosci.0822-10.2010>.
91. Elyse, H. Norton et al. “Human online adaptation to changes in prior probability”. In: *PLOS Computational Biology* 15.7 (July 2019). Ed. by Ulrik R. Beierholm, e1006681. issn: 1553-7358. <https://doi.org/10.1371/journal.pcbi.1006681>.
92. Edmund, Fantino, & Ali, Esfandiari. “Probability matching: Encouraging optimal responding in humans.” In: *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale* 56.1 (Mar. 2002), 58–63. issn: 1196-1961. <https://doi.org/10.1037/h0087385>.
93. Derek, J. Koehler, & Greta, James. “Probability matching and strategy availability”. In: *Memory & Cognition* 38.6 (Sept. 2010), 667–676. issn: 1532-5946. <https://doi.org/10.3758/mc.38.6.667>.
94. Nir, Vulkan. “An Economist’s Perspective on Probability Matching”. In: *Journal of Economic Surveys* 14.1 (Feb. 2000), 101–118. issn: 1467-6419. <https://doi.org/10.1111/1467-6419.00106>.
95. David, R. Wozny, Ulrik, R. Beierholm, & Ladan, Shams. “Probability Matching as a Computational Strategy Used in Perception”. In: *PLoS Computational Biology* 6.8 (Aug. 2010). Ed. by Laurence T. Maloney, e1000871. issn: 1553-7358. <https://doi.org/10.1371/journal.pcbi.1000871>.
96. Wolfgang, Gaissmaier & Lael, J. Schooler. “The smart potential behind probability matching”. In: *Cognition* 109.3 (Dec. 2008), 416–422. issn: 0010-0277. <https://doi.org/10.1016/j.cognition.2008.09.007>.
97. N. Tinbergen. “On aims and methods of Ethology”. In: *Zeitschrift für Tierpsychologie* 20.4 (Jan. 1963), 410–433. issn: 0044-3573. <https://doi.org/10.1111/j.1439-0310.1963.tb01161.x>.
98. Gerd, Gigerenzer, & Elke, M. Kurz. “Vicarious functioning reconsidered: A fast and frugal lens model”. In: *The Essential Brunswik: Beginnings, Explications, Applications*. Ed. by Kenneth R. Hammond and Thomas R. Stewart. New York, NY: Oxford University Press, 2001, 342–347. isbn: 9780195130133.
99. B. Wolf. *Brunswik’s Original Lens Model*. PDF, Brunswik Society / University of Landau. Retrieved from <https://brunswiksociety.org/wp-content/uploads/2022/06/WolfOriginalLens2005.pdf>. (2005).
100. Martin, T. Orne. “On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications.” In: *American Psychologist* 17.11 (Nov. 1962), 776–783. issn: 0003-066X. <https://doi.org/10.1037/h0043424>.

101. Ralph, Hertwig, & Andreas, Ortmann. “Experimental practices in economics: A methodological challenge for psychologists?” In: Behavioral and Brain Sciences 24.3 (June 2001), 383–403. issn: 1469-1825. <https://doi.org/10.1017/s0140525x01004149>.
102. Cass, R. Sunstein. “Nudges that fail”. en. In: Behavioural Public Policy 1.1 (May 2017), 4–25. issn: 2398-063X, 2398-0648. <https://doi.org/10.1017/bpp.2016.3>. https://www.cambridge.org/core/product/identifier/S2398063X16000038/type/journal_article (visited on 10/13/2021).
103. Aba, Szollosi, Nathan, Wang-Ly, & Ben R. Newell. “Nudges for people who think?”. In: Psychonomic Bulletin & Review (Jan. 2025). issn: 1531-5320. <https://doi.org/10.3758/s13423-024-02613-1>.
104. Ralph, Hertwig. “When to consider boosting: some rules for policy-makers”. In: Behavioural Public Policy 1.2 (Oct. 2017), 143–161. issn: 2398-0648. <https://doi.org/10.1017/bpp.2016.14>.

Acknowledgements

We thank Anna Dreber Almenberg, Ola Andersson, Magnus Johannesson, Erik Mohlin, Nurit Nobel, as well as seminar participants at the Arne Ryd workshop and the Swedish conference in economics.

Author contributions

BM and GKR collected and analysed data for the default and decoy conditions. MF and GKR collected and analysed data for the rule condition. MF wrote the manuscript based on a draft by BM with review and editing by GKR. All authors approved the final version of the manuscript.

Funding

Open access funding provided by Uppsala University. This work was supported by the Jan Wallander and Tom Hedelius Foundation, the Knut and Alice Wallenberg Research Foundation, the Swedish Research School of Management & IT, and the Marcus and Amalia Wallenberg Foundation.

Declarations

Competing interests

The authors declare no competing interests.

Consent to participate

All participants in this research gave informed consent before starting the experiment.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-32580-z>.

Correspondence and requests for materials should be addressed to M.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025