Neuro-Biological Origins of PT functionals*

Ferdinand M. Vieider[1]

[1]*RISLαβ, Department of Economics, Ghent University , ferdinand.vieider@ugent.be*

September 7, 2025

Abstract

Prospect theory (*PT*)—originally developed to accommodate challenges to expected utility theory such as the Allais paradoxes—has proven to be an exceptionally successful descriptive theory of choice. Nevertheless, the emergence of new paradoxes and a shift in "scientific taste" and interests over time have introduced some challenges to PT. I argue that the source of these challenges can be traced to the exogenous nature of the PT parameters, which has led to a proliferation of *ex post* explanations at the expense of *ex ante* predictions. In this chapter, I thus provide an overview of recent *generative* accounts of what I call "PT-like behaviour", which endogenize (subsets of) PT parameters based on neuro-biological and evolutionary principles. While still in their infancy, such models offer a perspective on where the stylised behavioural patterns discussed in the PT literature may originate.

Chapter prepared for the "Handbook of Prospect Theory", edited by Mohammed Abdellaoui and Han Bleichrodt.

*I gratefully acknowldge funding from the Research Foundation Flanders under the project "Causal Determinants of Preferences" (grant nr. G008021N). I am indepted to the following people for helpful comments and discussions about general preinciples, which in part pre-date the writing of this chapter: Rava Azeredo da Silveira, Aurélien Baillon, Jan Feld, Cary Frydman, Olivier L'Haridon, Nick Netzer, Ryan Oprea, Christian Ruff, and Peter P. Wakker. All errors and mis-interpretations are mine, and this chapter comes without any guarantee for accuracy. The opinions reflected are entirely my own, and do not reflect the opinions of the editors or the publisher.

1 Introduction

Prospect theory (PT) has proven very successful at describing observed choice patterns under risk—arguably its *raison d'être*. The focus on describing observed choices—resulting from its historical rationale of accommodating behavioural deviations from expected utility theory (EUT), such as the Allais paradoxes—has at the same time resulted in some limitations. These limitations become particularly apparent when examining what sort of behaviour is actually *predicted* by PT *ex ante*, as opposed to what it can explain *ex post*. This distinction is indeed crucial, given that *predictive* performance is generally taken to be a hallmark of scientific (i.e. empirically falsifiable) models.

Many of these limitations can be traced to the fact that PT is silent on the origin of the behavioural patterns described by its functionals. The parameters governing choices in the model are taken to be exogenous. This means that PT is not well suited to answer deeper questions about the origin of "preferences" or observed choice patterns (henceforth: *risk-taking*), which have received increasing interest over the years. This lack of insights on the origin of behaviour also shows when it comes to the *prediction* of behaviour across contexts. PT functionals have been shown to vary widely depending on the type of elicitation tasks (Hertwig, Barron, Weber and Erev, 2004; Feldman and Ferraro, 2023; Bouchouicha, Oprea, Vieider and Wu, 2024), yet no account of such context-dependence is possible within PT (or for that matter, any model applying stable preference functionals to objective choice primitives).

Here I review the increasing number of articles investigating the potential origins of the type of behaviour documented in the PT literature. I will refer to models detailing the deeper origins of behaviour as *generative* (as opposed to the descriptive–normative–prescriptive classification typically used in decision-making). The generative nature of the models thereby derives from their grounding in evolutionary or neural principles: they are supposed to pre-exist and causally determine choice. One essential question regards why we might be interested in such models at all. After all, revealed preferences are all economists typically care about. The grounding of most of these models in cognitive limitations and how we deal with them, however, has deep implications for our understanding of choice. Most disruptively, most of these models suggest that the choices we observe are not revelatory of preferences at all: Cognitive frictions in our perception or

manipulation of choice quantities, or difficulties we have mapping them into values, can result in choice behaviour that is to a large degree an expression of cognitive noise, rather than preferences. This does not imply that stable preferences do not exist: rather, they may be hidden by noise-induced reactions, and hence not revealed—a point to which I will revert in the discussion.

Nowhere does the exogeneity assumption show more than in the "source function" approach to ambiguity (Abdellaoui, Baillon, Placido and Wakker, 2011). Taken at face value, the approach could be used to organize violations of procedure invariance, simply by postulating that different elicitation methods constitute different "sources of risk". Given that such sources are identified *ex post* based on observed behaviour, the number of sources is potentially infinite. Temperatures in Rotterdam versus Tokyo are a popular example of different sources. This implies that weather in Paris, Antwerp, or The Hague would also be different sources. So would weather in Rotterdam next week. Or indeed objective probabilities arising from dice versus cards. One can see how this approach quickly spins out of control, de facto requiring an *infinity* of parameters. This is indeed a critique used against PT in general: its flexibility for *ex post* rationalization stands in the way of predictions, of which it arguably makes preciously few.[1]

Generative models hold the promise of counteracting this phenomenon. One way of seeing them is that they attempt to endogenize PT parameters by making testable predictions about how parameters governing the choice process ought to change under given circumstances. They thus hold the promise of explaining several empirical paradoxes arising under PT. One example are the opposite qualitative patterns of probability weighting functions when probabilities are described versus experienced (see Wulff, Mergenthaler-Canseco and Hertwig, 2018, for an overview). Others have emerged more recently: identical choices will solicit radically different behavioural patterns when presented in choice lists or as binary choices (Bouchouicha et al., 2024), or indeed when grouping choices into lists by probabilities versus sure outcomes (Feldman and Ferraro, 2023; Shubatt and Yang, 2024). Such patterns are incompatible with the application of pre-existing preference functionals to choice primitives, which are identical across contexts. Understanding

[1] Daniel Friedman (1989), who proposed a pioneering model showing how S-shaped utility could emerge from constrained optimization, describes exercises such as conducted in PT as follows: "Despite the high intellectual caliber of much of this work, there is an important sense in which it has been retrograde: theory is adjusted to evidence by weakening, not strengthening, its predictive power" (p. 1243).

where differences may come from will thus require a generative account endogenizing the model parameters.

Economic models such as subjective expected utility have long postulated overly strong rationality precepts (Bossaerts, Yadav and Murawski, 2019). PT—while being descriptive rather than normative in outlook—has maintained many of these strong principles: transitivity (shown to be violated already by Tversky, 1969), stochastic dominance, and strong separability precepts between probabilities and outcomes and between gains and losses, all of which are well-known to be violated (Birnbaum, 1999; 2008; Wu and Markle, 2008; Fehr-Duda, Bruhin, Epper and Schubert, 2010). Similar issues apply to rank-dependence, emphasizing mathematical coherence over the accurate description of behaviour. This has created a fundamental contradiction: while the focus of PT is purportedly descriptive, its emphasis on mathematical coherence in a deterministic setting was bound to result in violations of the model, given the fundamental stochasticity of behaviour known since the pioneering investigation of Mosteller and Nogee (1951). It should thus not come as a surprise that the generative accounts I present here challenge these specific aspects of PT, even while they are creating cognitive underpinnings that endogenize the model parameters.

PT is often defended by its proponents based on arguments of rationality and coherence. PT proponents have, for instance, been known to argue that the deterministic structure of the model makes it easy to handle, and that abiding by certain principles of mathematical coherence makes the model aesthetically pleasing. Unsurprisingly, these are the same arguments used by EUT apologists against PT. Taste, however, should not be mistaken for a scientific argument. It is also highly subjective: some of the generative models I discuss below implement constrained (stochastic) optimization as used in many fields in economics and beyond. The underlying mathematical and statistical principles are highly coherent, even though they result in violations of some of the strict rationality principles endeared to decision theorists. Elegance clearly lies in the eye of the beholder. Some of the generative models presented here would be considered "normative" when considered from a neuro-scientific and neuro-biological point of view. Given constraints in information processing, the *processes* being deployed to deal with these constraints will typically be optimal.[2]

[2]Criteria of optimality can of course be very different from economics. That simply occurs because

The point, however, runs deeper than this. One possible interpretative lens through which to see PT jointly with the historical context in which it evolved is the strength of the revealed preference paradigm (and the undeniable appeal of the underlying axioms). It is then not surprising that Kahneman and Tversky (1979) would have opted for a deterministic model, and chosen to ignore the elephant in the room that were the persistent choice inconsistencies. The implications of the models reviewed here is quite damning for revealed preferences. Virtually all of them—explicitly or implicitly—rely on some form of perceptual or combinatorial noise. To the extent that observed choice patterns may arise mainly from 'cognitive noise', one can no longer pretend that such choices can 'reveal' anything resembling a 'preference'. Pushing inconsistencies into a generic error term relegated to the basement of scientific interest neatly solved this issue, for which the economics profession anno 1979 would not have been ready. Even today, it raises profound issues about whether stable behaviour exists at all, and if yes, where it may originate—a question which will have to wait for the discussion to be re-examined.

This is not to say that generative models solve all of these issues. The approach is still relatively new, and as often happens in initial phases of discovery, some contradictions are starting to emerge between different generative models. While individual models are starting to be tested empirically—with some promising early evidence accumulating in their favour—comparative tests that would allow us to discriminate between different model families remain exceedingly rare. Another limitation is currently given by the sheer complexity of the neural functions which serve as an inspiration for most of these models—and more in general—by our still comparatively poor understanding of the functioning of the neural apparatus. This complexity means that existing models have largely focused on specific aspects of the underlying process that are more or less understood or that can easily be captured in the type of stylized models predilected by social scientists. For process models, however, details may matter more than for purely descriptive models—an issue that remains largely to be addressed.

In what follows, I will organize the chapter around the main stylized patterns in PT: decreasing sensitivity towards gains and losses, loss aversion, and probability weighting. Although not an integral part of PT, I also dedicate a separate section to stochastic

cognitive frictions may well make it impossible to directly maximize value or utility. Optimality will then be defined by the best possible strategy *conditional on cognitive noise*.

choice, which typically arises endogenously from the generative accounts I discuss. I will subsequently discuss the state of the empirical evidence, before circling back to a discussion of some of the more fundamental questions emerging from modelling generative versus descriptive patterns.

2 Decreasing sensitivity towards gains and losses

One of the hallmark patterns described in PT is decreasing sensitivity towards changes in wealth from a reference point—a principle that goes back at least to Markowitz (1952), and the foundational debate about the interpretation of expected utility theory (see e.g. Vickrey, 1945; Friedman and Savage, 1948).[3] Scoring changes in wealth on an absolute (and presumably invariant) utility scale, however, may be cognitively taxing. Tversky (1969) recognized this while discussing the superiority of comparative valuation when it comes to the ease of decisions. Nonetheless, Kahneman and Tversky (1979) decided to include utility transformation of absolute wealth changes into PT. Why—and possibly how—such valuations come about has received quite some attention. Here like elsewhere, I will discuss some select approaches more in depth at the expense of exhaustively reviewing the existing literature.

2.1 On the evolutionary origins of S-shaped utility

Why would people exhibit decreasing sensitivity towards changes in wealth? One answer may derive from the optimal—in the sense of biological fitness-maximizing—solution to a problem arising from limitations to the extent with which we can perceive and assign differences in value to consumption outcomes. Assessing the utility of consumption outcomes has instrumental value for the optimization of decisions. From an evolutionary design perspective, however, the more important question is how utility defined over per-period consumption can be exploited to maximize a long-term goal such as the maximization of evolutionary fitness (equated with the number of surviving offspring).

[3]Here as elsewhere, I will discuss PT as making predictions about behaviour (such as decreasing sensitivity twoards changes in wealth), in addition to providing mathematical functionals to summarize behaviour. While this interpretation may no longer be warranted—the fact that functions in PT can take *any* form is often emphasized by PT proponents when countering potential criticisms of the model—I believe that such an interpretation is coherent at least with early discussions in the PT literature, as well as being more compatible with the idea of PT being a scientific model that makes actual, falsifiable predictions, rather than a mere 'data summary apparatus'.

The answer provided by Robson (2001b) is that this requires a mapping from per-period consumption utility to utility defined over evolutionary fitness. Let x be evolutionary fitness, and let v be the utility of such fitness, with $v(x)$ the maximand in the problem. Given that x is a long-term outcome observable over the course of decades, and which therefore does not lend itself to immediate maximization, nature could plausibly have endowed humans with a more proximate utility function over consumption, call it $u(c)$. Since consumption is observed for each single period, adjustments to the consumption profile are more readily achievable. To ensure that consumption optimization lines up with fitness maximization, nature would also have created a mapping function from consumption to fitness. Let this mapping be $x = \psi(c)$. We can then directly define a utility function over consumption, $u(c)$, where $u \triangleq v \circ \psi$, i.e. $u(c) = v(\psi(c))$.

Robson (2001a) and Netzer (2009) base their predictions on the observation that individuals have limited cognitive capacity, which contrasts with potentially infinitely many values of c. Given the scarcity of neural resources, it will be evolutionarily optimal for an organism to allocate the finite number of perceptual thresholds at its disposal where they matter the most. The exact solution then depends on the optimization criterion. Robson (2001a) adopts the minimzation of the probability of mistakes as his criterion, and shows that in this case the utility ought to directly reflect the distribution of consumption opportunities in the environment (i.e. utility is equated with the CDF of consumption opportunities). Netzer (2009) argues that it will further be important to minimize the *size* of mistakes. Given that the size of mistakes will generally be inversely proportional to the probability of the mistakes, this implies that utility will not be quite as steep as implied by Robson's criterion, and that somewhat more "attention" will be dedicated to regions of consumption that are encountered less frequently.

The resulting utility curve is steepest in the regions corresponding to frequent outcome realizations, while flattening off in regions where consumption opportunities are less frequent. This results in an S-shaped utility function, and provides an evolutionary foundation for modelling decreasing sensitivity towards changes in wealth such as proposed by Markowitz (1952) and incorporated into PT by Kahneman and Tversky (1979). In this sense, marginal utility can be interpreted as the level of attention attributed to a consumption option. An additional feature of the model is that the peak of the empiri-

cal distribution functions of consumption opportunities endogenizes the reference point, which thus corresponds with the expected consumption level.[4] Other than the models we will see below, the Robson-Netzer framework thus endogenized the reference point in addition to utility curvature.

2.2 Comparative value and decreasing sensitivity

Mapping positive and negative changes in wealth onto an absolute (and presumably invariant) utility scale appears like an extremely taxing task from a cognitive point of view. One approach to making such valuations simpler is by comparing outcomes to 'typical outcomes' experienced by the decision maker. This results in comparative valuations as formally enshrined into the pioneering Decision-by-Sampling (DbS) model of Stewart, Chater and Brown (2006).

In the Decision-by-Sampling account, the value of an option will depend on its ordinal rank amongst outcomes experienced in the past. Concretely, any given outcome x to be evaluated is compared to a small number of samples drawn from memory. The utility of x will then be given by the relative rank R of x amongst the N drawn samples, $u(x) = \frac{R-1}{N-1}$. Stewart et al. (2006) show that the distribution of credits to UK bank accounts follows a power law, whereby moderately large payments are relatively frequent, whereas very small and very large payments are less frequent. This results in a power (constant relative risk aversion: $CRRA$) utility specification, providing micro-foundations for concave utility for positive changes in wealth. Stewart et al. (2006) furthermore document similar distributional patterns for debits (experienced losses), which gives rise to decreasing sensitivity towards losses (i.e. convex utility for losses).

The upshot of the model is that utility functions revealed by choices should reflect the experiences of the individual. Utility curvature is thus endogenously determined by the experiences of an individual. This also implies that the utility of a subject should change dynamically when exposing experimental subjects to different outcome ranges—a testable

[4] Glimcher and Tymula (2023) propose a neurally-based model of devisive normalization, whereby the payoff expectation drives the reference point. The model has the merit of including explicit dynamic equations for the payoff expectations, which are a weighted average of payoffs observed in the past, with weights decreasing as a function of temporal distance. The curvature of the function, in contrast, is treated as an exogenous parameter in the model, so that it is best seen as complementary to the approach presented here.

prediction to which I will return in the discussion about empirical tests.

2.3 Noisy Cognition and Bayesian Inference

A vast literature in neuroscience discusses an approximate number sense, which allows humans (as well as animals) to neurally code numbers semi-automatically on an analogue, approximate scale (see Dehaene, 2011, for a book-length discussion). Formal mathematical capabilities—while distinct and generally based on language processing—are furthermore thought to be linked to this approximate number sense. This suggests that, at least when many decisions are taken in short succession as typical for experiments, decision-makers may rely on approximate number assessments as inputs to their decision process. The question then becomes how to optimally deal with such noisy inputs—a question that is addressed in the noisy cognition model of Khaw, Li and Woodford (2021).

Khaw et al. (2021) (henceforth: *KLW*) use this principle to propose a model explaining small-stake risk aversion and decreasing sensitivity towards gains and losses (see also Woodford, 2012). Take a binary choice between a sure outcome c and a lottery paying a prize x with probability p, or else 0. KLW assume that outcomes c, x are perceived with noise, while p and 0 are perceived objectively. Noisy signals are modelled as single draws from distributions providing an unbiased representation of the logged outcomes: $r_x \sim \mathcal{N}(ln(x), \nu^2)$ and $r_c \sim \mathcal{N}(ln(c), \nu^2)$. The logarithmic mapping—which finds its justification in the logarithmic coding of numbers in the brain (Dehaene, 2003)—implies that the mental coding process implements Gustav Fechner's representation of Weber's law (Fechner, 1860; Thurstone, 1927b).

A simple choice rule may now consist in choosing directly based on this signal, as proposed by Thurstone (1927a). This results in a mechanism not dissimilar to the one in Decision-by-Sampling, although the soource of errors is technically different. The problem with this approach is that it is not optimal. Assume the DM has expectations about the probability distribution of $\{x, c\}$, enshrined in a prior distribution $ln(x), ln(c) \sim \mathcal{N}(\mu, \sigma^2)$. The DM can then calculate the posterior probability of the cause (the choice primitives x, c) of the noisy signals, which take the form $\mathbb{E}[ln(x)|r_x] = \frac{\sigma^2}{\sigma^2+\nu^2} r_x + \frac{\nu^2}{\sigma^2+\nu^2}\mu$ and $\mathbb{E}[ln(c)|r_x] = \frac{\sigma^2}{\sigma^2+\nu^2} r_c + \frac{\nu^2}{\sigma^2+\nu^2}\mu$. The signals will be taken into account in proportion to their precision, as measured by ν^{-2}: the less precise the signal, the less the DM will rely on it. While

this approach introduces systematic bias in the form of regression to the prior mean, it is optimal inasmuch as it minimizes the mean squared error of the estimator (see Bishop, 2006, chapter 3). Incorporating these inferences into a choice rule aiming to maximize expected value yields:

$$Pr[(x,p) \succ c] = \Phi\left[\frac{\alpha\, ln(x/c) - ln(p)^{-1}}{\sqrt{2}\alpha\nu}\right],$$

where Φ represents the standard normal CDF (making this a Probit model), and $\alpha \triangleq \frac{\sigma^2}{\sigma^2+\nu^2}$. Choice will be inherently stochastic because of the randomness in the signals r_x, r_c. Focusing on the numerator and exponentiating, we see that the choice likelihood will be proportional to $px^\alpha - c^\alpha$, thus providing cognitive micro-foundations for as-if CRRA utility. This is indeed how the model explains small-stake risk aversion—assuming that ν is independent of $\{x, c\}$, apparent risk aversion inferred from behaviour will be dissociated from the stakes involved.[5]

The model provides an excellent illustration of how optimal ways of dealing with cognitive frictions can produce as-if utility functions. Applied separately to gains and losses (with the sign assumed to be perceived perfectly), the model provides cognitive micro-foundations to decreasing sensitivity towards changes in wealth. The degree of decreasing sensitivity is thereby solely driven by coding noise: a DM who perceives numbers perfectly and without noise will simply maximize expected value (although expected utility maximization over large stakes could enter the model through the decreasing marginal utility of wealth). An interesting question may thus concern the drivers of noise, and to what extent noise may be an individual characteristic versus a contextual effect. I will return to such more general questions in the discussion.

2.4 Sensitivity towards Costs and Benefits

The noisy cognition model of Khaw et al. (2021) micro-founds coding noise on noisy number perception. That is, however, not the only way of introducing errors. In Viei-

[5] An interesting question concerns what might happen over large stake variations. The typical empirical observation is *increasing* relative risk aversion as stakes are scaled up (Holt and Laury, 2002; Fehr-Duda et al., 2010; Bouchouicha and Vieider, 2017). Under the KLW model, this obtains in the context where coding noise ν *increases* in stakes, so that α decreases in stakes. Enke and Shubatt (2023) document that subjects indeed a) perceive high-stakes choices as more difficult compared to low-stake choices; and b) make more mistakes for such high-stake choices.

der (2024), I develop a model of probability weighting based on signal detection theory (Green, Swets et al., 1966). For 50-50 lotteries, probability weighting drops out and the model provides micro-foundations for errors in the perception of costs and benefits of the lottery relative to the sure outcome. Defining errors over costs and benefits has an important pedigree in neuroscience, being grounded in log-odds coding of favourable versus unfavourable information (Gold and Shadlen, 2001). Errors occurring at the stage of the recombination of choice quantities are furthermore thought to be more important than purely perceptual errors, given much of the cognitive bottleneck affecting our processing of real-world information seems to stem from higher-order cognitive processes (Drugowitsch, Wyart, Devauchelle and Koechlin, 2016; Zheng and Meister, 2024).

Take a 50-50 lottery paying x or else y, and compare this to a sure outcome of c. A DM wanting to maximize expected value will choose the lottery whenever $p\,(x-c) > (1-p)(c-y)$ (obtained by decomposing $c = p\,c + (1-p)\,c$), i.e. whenever the expected benefits of taking the lottery exceed the expected costs. Assuming that costs and benefits are subject to noisy signals and that $p = 0.5$, the choice rule will entail choosing the lottery whenever $\mathbb{E}[ln(x-c)|r_b] > \mathbb{E}[ln(c-y)|r_k]$, ,where the subscripts to the signals stand for costs and benefits, respectively. The derivation then proceeds just like above, resulting in a stochastic choice rule entailing decreasing sensitivity towards costs and benefits. Describing the noisy signal as directly affecting the log cost-benefits seems efficient in this context, while being inconsequential from the point of view of empirical predictions. Such a signal for the ratio, indeed, fits the influential idea in neuroscience that there may be groups of neurons signalling positive aspects of a decision, and group of "anti-neurons" signalling the negative aspectes (Gold and Shadlen, 2001; 2002).

Note that errors on single outcomes and errors on costs and benefits may well co-exist, and one could easily build a more general model combing cost-benefit distortions with single number distortions as in KLW. Tversky (1969) provides an early discussion of the generality of comparative setups defined over relative benefits as shown here, since they nest the different sub-cases described above. Nonetheless, the model deviates from decreasing sensitivity in PT—and if taken literally—suggests that PT may be mis-specified in this respect. Here, the discussion serves mainly as a setup for the later discussion of probability weighting. It is nevertheless worth noting that the intuition of comparative

evaluations facilitating the decision process presents some parallels with the models discussed above, notably with the comparative mechanism underlying DbS, even though the technical details are different.[6]

3 Attitudes towards mixed gain-loss lotteries

Somewhat surprisingly, given the importance of the concept of loss aversion in the behavioural economics literature, generative accounts for loss aversion appear to be somewhat thinner on the ground than for either decreasing sensitivity or probability weighting. One reason may be the added complexity arising from mixed domain choices. Here, I will focus on three main accounts. As usual, I will discuss the general implications of the models, at times going beyond the specific mechanisms discussed within the papers proposing them.

3.1 Loss aversion in Robson-Netzer

The Robson-Netzer framework naturally endogenizes the reference point by identifing it with the peak of the probability density function of consumption opportunities. It does, however, not automatically produce a discontinuous kink at that point, nor should one generally expect it to be asymmetric. Netzer (2009) thus discusses a separate mechanism that could produce such a kink in utility, which passes through a model of how time delays correlating with the size of consumption may impact choices.

Although not explicitly modelled by Netzer (2009), there is arguably an alternative mechanism in the model by which loss aversion could enter the utility function. Given the mapping between consumption and evolutionary fitness discussed above, Netzer (2009) shows that consumption utility ought to result from a weighted average of the CDF of consumption opportunities, $F(c)$, and the fitness function, $\psi(c)$, mapping consumption into evolutionary fitness. Intuitively, attention should be allocated to where mistakes are likely to be most frequent, but also to where they may be most impactful in terms of evolutionary long-term goals, as captured by the mapping ψ.

[6] An open question concerns the precise mechanism underlying the Bayesian combination of evidence and prior. One possibility is indeed that this combination happens by combining the samples representing the coding function with samples from memory, which would indeed suggest a combination of Bayesian inference processes with a sampling mechanism such as modelled in DbS.

This opens a new window on how particular attitudes towards losses could arise in the model. It seems indeed plausible that, when measured against the reference point of expected consumption, temporary shortfalls in consumption would have a greater impact on evolutionary fitness than windfall gains. This conclusion follows directly from the observation that a stable consumption profile is necessary to guarantee the survival of any offspring. It seems thus plausible that the mapping function ψ would be asymmetric around expected consumption. It is less clear that this would result in a kink at the reference point, not least because biologically grounded processes rarely present such discontinuous changes. A plausible mechanism could be that attention to losses should increase with the size of the losses, thus resulting in more gradual changes in loss attitudes as already discussed by Markowitz (1952). Similar intuitions will indeed emerge from the models I review next, even though they have quite different starting points.

3.2 Decision-by-Sampling and loss attitudes

Other than in the Robson-Netzer framework discussed above, the DbS model is formulated over gains and losses, which are assumed to be perfectly distinguishable. Other than in PT, however, the model does not generate a dedicated parameter governing choices amongst mixed gambles: loss attitudes instead emerge from differential utility curvature for gains versus losses.

Utility in DbS is a product of the relative ranking of outcomes when compared to a handful of draws from memory. Stewart et al. (2006) show that while gains (credits to bank accounts) and losses (debits to bank accounts) both follow a power law, the distribution for losses is shifted closer to 0. That is, few relatively large credits (such as salary payments or gifts) are usually put towards many small expenses (such as grocery shopping or bills). The upshot is that a given gain x will usually receive a lower rank when compared to typical gains than an identical loss x when compared to typical losses. This will result in sensitivity to a given loss x generally being higher than sensitivity to a monetarily equivalent gain x. DbS thus predicts dislike of mean-preserving spreads around 0, although it does so without a dedicated parameter, and purely based on differential sensitivity towards gains and losses. In addition to generally risk averse choices, we may thus expect that the degree of risk aversion will increase with stake size—something that has indeed been documented in the empirical literature (see e.g. Ert and Erev,

2013).

Note that PT can of course account for the same type of stake dependence if sensitivity towards negative changes in wealth is more pronounced than sensitivity towards positive changes, so that stake effects per se do not contradict PT. Given that parameters are taken as exogenous and "purely descriptive", however, it does not ex ante *predict* such an effect. The meta-analysis of PT parameters by Imai, Nunnari, Wu and Vieider (2025), however, lends support to the hypothesis of heightened sensitivity towards losses by showing that utility for losses is typically closer to linearity than utility for gains. Bouchouicha, Li and Vieider (2025) (who conduct one of several empirical tests of DbS, reviewed below) thus coined the term loss-*sensitivity* to describe this specific mechanism.

3.3 Noisy Cognition and Attitudes towards Losses

Somewhat uniquely amongst the neuro-biologically founded models discussed here, the noisy cognition model of Khaw et al. (2021) provides cognitive micro-foundations for a separately defined kink in the utility function at the origin. Just like in DbS, that origin is identified with the status quo of (nominal) 0, since gains and losses are assumed to be perfectly discriminated. The model can thus account for loss aversion in the PT sense of a loss aversion parameter $\lambda > 1$, while at the same time detailing under what circumstances such loss aversion ought to be observed. It does, however, also incorporate a separate mechanism for loss-sensitivity, thus approaching PT in its richness.

Take a 50-50 gamble offering a gain G or else a loss L as a running example. Other than for the general case described above, the model of KLW and my own model (Vieider, 2024) coincide for this special case. Gains and losses will be subject to noisy neural encoding, so that $r_g \sim \mathcal{N}(ln(G), \nu_g^2)$ and $r_\ell \sim \mathcal{N}(ln(L), \nu_\ell^2)$. In the model's most general version, the two signals will then be decoded with separate priors for gains and losses: $ln(G) \sim \mathcal{N}(\mu_g, \sigma_g^2)$ and $ln(L) \sim \mathcal{N}(\mu_\ell, \sigma_\ell^2)$. The stochastic choice rule will thus look as follows:

$$Pr[(G, 0.5; L) \succ 0] = \Phi \left[\frac{\alpha\, ln(G) - \beta\, ln(L) + (1-\alpha)\, \mu_g - (1-\beta)\, \mu_\ell}{\sqrt{\nu_g^2 \alpha^2 + \nu_\ell^2 \beta^2}} \right]. \qquad (1)$$

where Φ as usual represents the standard normal CDF, and where $\alpha \triangleq \frac{\sigma_g^2}{\sigma_g^2 + \nu_g^2}$ and $\beta \triangleq$

$\frac{\sigma_\ell^2}{\sigma_\ell^2+\nu_\ell^2}$ can be thought of as discriminability parameters for gains and losses, respectively capturing how accurately the true causes (the gains and losses constituting the choice primitives) are reflected in choice.[7]

To illustrate how the model can produce a kink in utility, let $\lambda \triangleq exp[(1-\beta)\mu_\ell - (1-\alpha)\mu_g]$. Substituting $ln(\lambda)$ into the numerator and exponentiating it, one can see that acceptance proportions of the gamble are predicted to be proportional to $G^\alpha - \lambda L^\beta$. This provides cognitive micro-foundations for loss aversion as modelled in PT. Loss aversion in the sense of $\lambda > 1$ will then occur whenever $(1-\beta)\mu_\ell > (1-\alpha)\mu_g$. Loss aversion thus requires expected losses to exceed expected gains.[8] Although such a motive could have been hardwired by nature, the optimality of such a mechanism in evolutionary terms is questionable. For one, gains tend to be larger than losses on average both in experiments and in the real world (think about the levels of growth in per capita income humans have experienced over the last 400 years or so). Hardwiring a prior expectation may further limit learning, and thereby inhibit the adaptive mechanism that is the main tool provided in the model to overcome limitations arising from biological constraints to information processing.

The model also enshrines cognitive micro-foundations for loss-*sensitivity*. Loss sensitivity here coincides with $\beta > \alpha$, which will occur whenever normalized coding noise for losses is lower than for gains, i.e. whenever $\frac{\nu_\ell}{\sigma_\ell} < \frac{\nu_g}{\sigma_g}$. Bouchouicha et al. (2025) show theoretically that this prediction obtains endogenously from the model whenever more attention is given to losses than to gains, whereby attention is simply equated with dwelling time on a particular attribute. More attention to losses than gains in mixed choices is indeed supported by experiments measuring dwelling times (Pachur, Schulte-Mecklenbeck, Murphy and Hertwig, 2018) or using eye-tracking techniques (Hirmas, Engelmann and Weele, 2024). The model furthermore provides a new lens through which to examine the celebrated results of Tom, Fox, Trepel and Poldrack (2007). Combining behavioural with brain-scanning data, the latter showed that negative neural activity decreases more markedly with the size of the loss than neural activations increase with the size of the gain.

[7] I here use the means of the priors, thereby slightly deviating from the derivation in Khaw et al. (2021). The difference arises from first logging the choice rule, and here serves mainly to simplify the formal setup.

[8] What I mean by this is that the expectation incorporated in the prior for losses exceeds that for gains, $\mu_\ell > \mu_g$. This condition rests on the additional assumtion that—a least for most individuals—$(1-\beta) \leq (1-\alpha)$. I will discuss the rationale underlying this assumption shortly.

Neural activation differentials were further highly correlated with acceptance decisions of mixed gambles. These results thus lend direct support to the notion of loss-sensitivity. The discussion of positive versus negative activations further literally fits the idea of neurons versus anti-neurons proposed by Gold and Shadlen (2001).

The intuition of "increased attention to losses" thus emerges as a common element from the three modelling approaches discussed in this section, even though it emerges from rather different mechanisms. This shows how evolutionary, neural coding, and statistical frequency approaches are much more similar to each other than a superficial categorization may suggest. As we will see when discussing empirical tests, the subtle differences in how the models are written will nevertheless allow to empirically discriminate between at least some of these models.

4 On the origin of probability distortions

We have now examined several models that provide cognitive and evolutionary microfoundations for behaviour triggered by changes in wealth. Several of these models also have extensions specifically targeting probability distortions. Here, I will as usual examine a subset of these models, while electing to discuss this subset in more depth.

4.1 Probability weighting as a second-best

Probability distortions are not directly included in the Robson-Netzer framework discussed above. Herold and Netzer (2023), however, present a model where they characterize inverse S-shaped probability weighting as an optimal reaction to S-shaped utility. In this sense, probability-distortions are a "second-best" solution that serves to compensate for distorted outcome perceptions (see also Steiner and Stewart, 2016, for the notion of probability distortions as a second-best deriving from optimally distorted signals).

Here, I focus on the characterization of probability weighting in Netzer, Robson, Steiner and Kocourek (2024). Uniquely amongst the models examined here—which generally make their case based on simple tradeoffs between binary lotteries under risk—Netzer et al. (2024) specifically target complex multi-outcome lotteries under uncertainty. They characterize biased decisions emerging from a combination of frictions in the encoding (or

measurement) and the decoding (or aggregation) process. A DM will first of all measure the rewards provided in a given situation. Given a fixed precision in overall signals, this problem amounts to distributing the signal precision across different states, resulting in implications similar to those discussed above for decreasing sensitivity.

The aggregation or decoding stage is just as important. For instance, a DM may bundle several states that she sees as somehow similar into one partition of the reward space, thereby simplifying the decision problem. Even for a risk neutral decision maker, such bundling will inevitably result in behavioural deviations from expected value maximization. For instance, if several extreme outcome scenarios—which will typically have relatively small probabilities attached to them—are bundled into one partition and attributed the most extreme reward, then DMs will oversample from that partition, i.e. allocate too much attention to it. The consequence will then be that such extreme events are overweighted relative to more commonly observed events, yielding the canonical inverse S-shaped probability distortions.[9]

The model of Netzer et al. (2024) ultimately unifies and generalizes previous accounts in the same model family. When the encoding function is allowed to optimally adapt to the distribution of rewards (in an ancestral environment, given the evolutionary motivation of the model), the outcome distortions and probability distortions are jointly optimal. The optimality of probability distortions, in particular, rests on the argument that tail events offering very high or very low rewards typically occur with small probabilities. Given that the reward-coding strategy is affected by relatively high noise for such tail events, a DM will often have difficulties reaching decisions involving lotteries with such extreme events. This, in turn, leads to oversampling for such extreme events and to probability distortions, which thus compensate for noisiness in reward representations. This mechanism endogenizes probability distortions as a function of reward, which could potentially account for findings according to which probability distortions become more extreme as stakes increase (Rottenstreich and Hsee, 2001; Bouchouicha and Vieider, 2017).

[9]The examples in Netzer et al. (2024) typically use real world scenarios, and effects of sampling are presented as "subjective probabilities" deviating from "objective probabilities". From a PT perspective, this may seem as a mechanism of belief formation, rather than a mechanism of probability distortion. Note, however, that from the perspective of this model—and indeed from the perspective of most of the models presented here—this distinction is not meaningful, since even objectively given and described probabilities are subjective "beliefs" given their noisy perception. I will return to this point in the discussion.

4.2 Likelihood frequency and likelihood distortions

DbS contains a mechanism that attributes inverse S-shaped probability distortions to the distribution of probabilities experienced in the environment. In particular, Stewart et al. (2006) argue that—in several contexts—small and large probabilities are more frequent than intermediate probabilities. This, in turn, implies that small probabilities and large probabilities will, on average, be ranked more accurately than intermediate probabilities. By the same arguments used above, such distributions would then entail inverse S-shaped probability distortions.

Frydman and Jin (2023) propose a more sophisticated version of the same argument. In their model, learned frequency distributions of probabilities enshrined in a Bayesian prior serve to optimally adapt noise in probability perceptions to the expected circumstances. Basing themselves on the efficient coding model of Heng, Woodford and Polania (2020), coding noise should then be larger over intermediate ranges than for extreme probabilities if DMs have a U-shaped prior. This, in turn, will result in likelihood-insensitivity over intermediate probability ranges. Just like Decision-by-Sampling, the model thus predicts that probability weighting should change as a function of the frequency with which small and large versus intermediate probabilities are observed and with the frequency of small versus large probabilities.

The predictions are interesting for a number of reasons. For one, reducing coding noise for small and large probabilities in the binomial setup used for the model implies that encoding very small and very large probabilities precisely will be very expensive in terms of cognitive resources required due to the skewed nature of the binomial (or Beta) distribution. Given the fixed cognitive resources setup they inherit from Heng et al. (2020), the U-shape in the prior ought to be quite accentuated to make such a strategy optimal. The prediction is exactly the opposite of the one by Zhang, Ren and Maloney (2020). The latter model a fixed signal bandwidth, which results in extreme probabilities being shrunk towards the end points of that bandwidth, and thus treated as less extreme than they truly are in choice. Probability distortions then result from extreme noise affecting the smallest and largest probabilities—the exact opposite of the prediction by Frydman and Jin (2023).

4.3 Probability Distortions as Optimal Bayesian Inference

A popular probability weighting function due to Goldstein and Einhorn (1987) takes the form $w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}$. This functional form has a particularly intuitive interpretation in its linear in log-odds (LLO) form, characterized by Gonzalez and Wu (1999):

$$ln\left(\frac{w(p)}{1-w(p)}\right) = \gamma\, ln\left(\frac{p}{1-p}\right) + ln(\delta). \tag{2}$$

In this form, values of $\gamma < 1$, which takes the form of a power to the odds, can directly be seen to compress the odds towards 1 (with $\gamma > 1$ having the opposite effect). The intercept term $ln(\delta)$ further modifies the fixed point towards which the log-odds are compressed (i.e. the elevation of the log-odds function when the log-odds are 0, i.e. for $p = 0.5$). Zhang and Maloney (2012) provide an early discussion of how such a function can obtain from a weighted combination of the stimulus and a Bayesian prior mean.[10]

In Vieider (2024), I propose a Bayesian Inference Model (BIM) of probability weighting that builds on the Bayesian inference framework used by Khaw et al. (2021), but that is strictly seen not a generalization of that model because of outcome distortions being defined over costs and benefits as detailed above. An important characteristic is that the model provides an integrated perspective in which both probabilities and outcomes (costs and benefits) are noisily perceived. A further distinguishing characteristic of the model is that—other than the models reviewed previously in this section—this model is geared explicitly towards binary choice. The reason for this is that, under the Bayesian Inference perspective, binary choices and choice lists or valuation tasks require separate models. This is the true sense in which these are process models: while the neuro-cognitive process used across different contexts remains the same, the way this process is applied depends fundamentally on the characteristics of the choice situation.[11]

Take a simple tradeoff between a binary lottery $(x, p; y)$ and a sure outcome c. Using the optimal choice rule already discussed above generalized to any probability p, I assume

[10] Zhang and Maloney (2012) show empirical how such log-odds representations are pervasive, providing a good empirical fit not only to decision patterns under risk and uncertainty, but also to belief formation in a variety of contexts (see also Enke and Graeber, 2023). Zhang et al. (2020) augment this setup by a stylized noisy signal model, where probabilities in intermediate ranges are coded perfectly, but extreme probabilities drop out of the signal space and are thus perceived as less extreme than they truly are.

[11] Khaw, Li and Woodford (2023) and Bouchouicha et al. (2024) independently and simultaneously developed models of probability weighting in valuation tasks, to be further discussed in the review of empirical predictions and tests below.

that the mind trades off the log-odds against the log cost-benefits. Just like for the cost-benefits, the log-odds are noisily coded by a signal $r_p \sim \mathcal{N}(ln\left(\frac{p}{1-p}\right), \nu_p^2)$. Decoding this signal by combination with a Bayesian prior $ln(\frac{p}{1-p}) \sim \mathcal{N}(ln(\eta), \sigma^2)$, and taking the expectation over many repetitions of an identical probability p, yields the average observable response to the choice primitive p:

$$\mathbb{E}\left[\mathbb{E}\left[ln\left(\frac{p}{1-p}\right) \mid r_p\right] \mid p\right] = \frac{\sigma^2}{\nu_p^2+\sigma^2} ln\left(\frac{p}{1-p}\right) + \frac{\nu_p^2}{\nu_p^2+\sigma^2} ln(\eta). \tag{3}$$

Defining $\gamma \triangleq \frac{\sigma^2}{\nu_p^2+\sigma^2}$ and $\delta \triangleq \eta^{1-\gamma}$ yields a function identical to the LLO probability weighting function from which we started. Likelihood-insensitivity here has a cognitive interpretation, and is driven by the noise in probability perception (possibly capturing an imprecise notion of what a probability of e.g. 1/8 truly means, an intuition that can be tested). To reach a decision, this inference can now be stochastically traded off against the cost-benefit perceptions derived above.

It is important to spell out that this is not a "just-so-story" of what *might* underlie probability distortions. The model gives a precise characterization of probability weighting in terms of cognitive noise. In Oprea and Vieider (2024), we adapt the model to a sampling context and show that it explains the description-experience gap. The starting point is to characterize the rather abstract mental signal r_p, and to obtain a parallel notion in a setting where probabilities need to be learned by sampling from unknown options. Assume that subjects are sampling from a binary lottery. Let $\alpha = \sum_{i=1}^{N} s_i$ be the count of successes (draws of the prize x) in N draws from the lottery, and $\beta = \sum_{i=1}^{N}(1-s_i)$ the count of failures (draws of failures y). We can then characterize probabilistic beliefs about p at any given stage as being described by a Beta distribution parameterized by the count of successes and failures, $\mathcal{B}e(\alpha, \beta)$. Using the logit-normal distribution derived by Atchison and Shen (1980), the mean in log-odds space will be given by $ln(\frac{\alpha}{\beta})$, and noise will be given by $\nu_p^2 = \digamma'(\alpha) + \digamma'(\beta)$, where $\digamma'$ is the trigamma function. Note that the sampling framework can also be applied to description-based choice. The only difference is that parameters like α and β will now characterize "mental samples", which can be thought of as neuronal firing rates or action potentials.

4.4 Explaining the Ellsberg paradox

The explanations proposed above were specifically tailored to explaining behaviour under risk. It is, however, straightforward to generalize the Bayesian inference setup above to explain ambiguity attitudes and the Ellsberg (1961) paradox. I will briefly illustrate this for the Ellsberg 2-colour problem following the model of L'Haridon, Oprea, Polania and Vieider (2023).

The model builds on the sampling framework introduced above. Let the ambiguous probability be encoded by $\mathcal{B}e(\alpha_a, \beta_a)$ and the risky probability by $\mathcal{B}e(\alpha_r, \beta_r)$. Following arguments of colour exchangeability as presented by Raiffa (1961), let us assume that decision-makers actually recognize the mean probability in the two urns to be the same, so that $\frac{\alpha_a}{\alpha_a+\beta_a} = \frac{\alpha_r}{\alpha_r+\beta_r}$. If DMs are less certain of their assessment of the ambiguous probability, in the sense that the concentration of the distribution is smaller for the ambiguous urn, i.e. $\alpha_a + \beta_a < \alpha_r + \beta_r$, this will result in increased coding noise for the ambiguous option relative to the risky, i.e. $\nu_a > \nu_r$. Further parsimoniously assuming a *common* prior $\mathcal{N}(\mu, \sigma^2)$ yields the following stochastic choice equation:

$$Pr[p_r(x) \succ p_a(x)] = \Phi \left[\frac{(\rho - \gamma) \times \left[ln\left(\frac{p}{1-p}\right) - \mu \right]}{\sqrt{\lambda_r^{-1}\,\rho^2 + \lambda_a^{-1}\,\gamma^2}} \right], \tag{4}$$

where ρ and γ are the Bayesian evidence weights for risk and ambiguity, and a defined as $\rho \triangleq \frac{\sigma^2}{\sigma^2+\nu_r^2}$, and $\gamma \triangleq \frac{\sigma^2}{\sigma^2+\nu_a^2}$.

The model has a few remarkable characteristics. Starting from an intuition not unlike the one underlying multiple-prior models (Gilboa, 1987; Ghirardato and Marinacci, 2002), ambiguity aversion obtains from risk aversion in the prior, captured by $\mu < 0$. At the same time, however, the reduced confidence in the ambiguous probability results in likelihood-dependence of ambiguity attitudes much as documented in PT. Under risk, the model naturally falls back to probability distortions generated by less-than-perfect coding of objective probabilities, which will occur as long as $\alpha_r, \beta_r < \infty$. Risk is thus not not an end-point in the continuum of degrees of uncertainty, but may well occupy an intermediate position.

5 Choice stochasticity

PT is a deterministic model: if one can measure the preferences of a DM, one ought to also be able to predict all future choices of that DM. Although an error model needs to be attached to PT to empirically recover its parameters, such error models have received little attention in the PT literature. Error models added to PT have thus typically been chosen independently, often to reflect zero-mean 'white noise' processes. Modelling errors as completely independent of the decision process may, however, produce complications, as detailed e.g. by Wilcox (2011) and Apesteguia and Ballester (2018). The details here may further depend on the choice setup used for the measurement—another issue on which PT is silent, given its universalist aspirations.

Almost all of the generative accounts discussed above (with the exception of Glimcher and Tymula, 2023) contain some mechanism endogenizing noise. Even though the specific aspects from where noise arises differ between models, a prediction emerging from all models is that noise will generally be more important in regions of stimuli that are less frequently encountered. Here, I present a brief overview of how errors arise in the main modelling paradigms reviewed above.

Errors in Robson-Netzer

Errors in Robson-Netzer arise from the distribution of the thresholds at which jumps in just-noticeable-differences in utility become detectable. This happens because utility takes the form of a step function (even though the number of steps can potentially be very large). Intuitively, any two consumption levels falling between two subsequent steps receive identical utility. The choice between them is thus predicted to be random.

Given that differences between consumption thresholds for which changes in utility are detected will be an increasing function of the distance to the peak of the probability density function of consumption, errors will be larger for rarely-encountered consumption ranges. A criticism that has been voiced about the formal model setting is that small increments in consumption will be undetectable when they fall into the interior of the interval between two threshold, but identical changes in consumption might cause jumps if they are added to consumption levels close to a threshold. Such an interpretation, however, may be too literal in a stylized modelling setting such as this. The fact that

most results are obtained in the limit as the number of thresholds becomes infinitely large furthermore suggests that this problem may be minor in practice.

Sampling-based errors

Just like consumption opportunities in the Robson-Netzer framework, monetary gains and losses in DbS are perfectly perceived. Noise will, however, arise in the process of utility assignment. In particular, the source of noise now stems from the low number of samples taken from the distribution of outcomes in memory, which results in the rank of the outcomes to be assessed being measured with low fidelity. Formally, the errors in the model take the form of a draw from a binomial distribution, and are directly attached to the utility of the outcomes, thus resulting in a random utility model.

While the error term naturally arises from the sampling process (and is in this sense endogenous to the model), it is assumed to be independent from the outcomes to be assessed. The implication of this is that, once more, the error frequency should increase for outcomes falling far from the typical distribution to which a decision maker is adapted, regardless of whether they are smaller or larger. This happens because extreme outcomes that fall relatively close together in the distribution are likely to receive similar rankings. The error term is then more likely to overwrite any differences in ranking than would happen for outcomes falling into the middle of the habituated distribution, where rank assignment will be more accurate. At the extreme, choices between unusual quantities may thus converge towards almost random behaviour.

Note that this provides cognitive foundations for the behaviour of the random utility model characterized by Apesteguia and Ballester (2018). Other than in their characterization, where the resulting stochastic non-monotonicities are an undesirable feature of a statistical estimator, in DbS they are a natural corollary of difficulties in assigning utilities to outcomes. Whereas for very small ranks (strong concavity of utility) the predictions of the two approaches coincide, DbS makes similar predictions for unusually large outcomes, where choice should once again converge towards randomness. This is a prediction that is not shared by the purely statistical characterization of Apesteguia and Ballester (2018), thus providing a potential test for the stochastic choice predictions of DbS.

Coding noise versus decision noise in Bayesian inference

Bayesian inference requires a more nuanced discussion of errors, not least because there are different types of errors in the model. The first category of errors is simply given by coding noise—the inverse of the precision with which choice primitives are neurally encoded. As a general rule, regression to the mean will increase in the weighted distance of a stimulus to the expected stimulus, given that Bayesian shrinkage is a function of, *inter alia*, the *prediction error*, $x-\mu$. In addition, coding noise itself should be expected to adapt to the distribution of stimuli expected in the environment—a phenomenon referred to as *efficient coding* (see e.g. Heng et al., 2020, for a model of efficient coding in the context of noisy perception; see Frydman and Jin, 2022, for an application). While only a subset of the models reviewed above represent noise adaptation explicitly, such adaptation can nevertheless be seen as a general feature of these models. For instance, while Khaw et al. (2021) discuss as-if utility incorporating constant relative risk aversion, allowing for coding noise to increase for larger (and less commonly experienced) numerical stakes results in a prediction of *increasing* relative risk aversion.

Some complications, however, may arise in other model specifications. For instance, in the probability weighting model of Vieider (2024), regression to the mean of the Bayesian prior will depend on the interplay between coding noise for probabilities and coding noise for cost-benefits (where coding noise is to be understood as measured relative to the variance of the prior). In a dynamic context where coding noise may adapt, this raises a number of potentially interesting but so far unaddressed questions of how the different types of coding noise will adapt, under which circumstances attention may shift from one dimension to the other, and whether and to what extent existing correlations in choice stimuli may be correctly reflected in the correlations of noisy signals (see e.g. Natenzon, 2019, on a model exploiting noise correlations in a different context).

The Bayesian inference model is unique amongst the models discussed here in that it does not predict an unqualified increase in *decision noise* (i.e. inconsistencies actually observed in choices) for stimuli falling far from the expected range. The reason for this can be found in the intuition underlying the optimality of the Bayesian estimator, which entails that decision noise is a non-monotonic function of coding noise. This can most easily be shown in the one-dimensional model of Khaw et al. (2021). For that particular model,

the maximimal decison noise will be observed when the ratio of the SDs of the coding noise parameters of the prior is equal to 1, i.e. for $\frac{\nu}{\sigma} = 1$. For values $\nu < \sigma$, inferences are fairly accurate, and trial-to-trial variation resulting from draws of independent signals from the likelihood will thus be less important. For values $\nu > \sigma$, however, the best reaction to the large coding noise is to increasingly rely on the prior in decoding, which again will reduce decision noise and result in more consistent behaviour. The upshot of this is that the models of Khaw et al. (2021) and Vieider (2024) do not present the non-monotonicities in stochastic choice characterized by Apesteguia and Ballester (2018): as-if utility and probability weighting are tightly linked to the stochastic model, thus resulting in a monotonic stochastic choice function.

6 The state of the empirical evidence

Even while providing cognitive micro-foundations for PT-like behaviour, many of the models presented in this chapter make distinctive predictions that can be used to distinguish them from PT. That being said, not all models are created equal, in that the predictions of some are more specific, and hence more testable, than others. This is indeed natural for models formulated at different levels of abstraction. It does, however, make it more difficult to distinguish between predictions that are common to the models, and predictions that could help discriminate between the generative models themselves. Here, I will attempt a review of the empirical evidence, with particular attention to elements separating the models from PT, and from each other.

6.1 General tests of cognitive frictions

Several papers have tested general implications of noisy cognition. Enke and Graeber (2023) show probability distortions, conceptualized as regression to the mean, to correlate with answers to a survey question asking subjects how certain they are about their choice (see also Enke, Graeber, Oprea and Yang, 2024, for a paper showing that similar "behavioural attenuation" is at work across a large variety of contexts). Oprea (2024) presents a setup in which rewards are described by the contents of 100 boxes. A lottery is then represented by a random draw from 100 boxes containing different rewards. He then proceeds to creating a situation in which the complexity of the choice situation is

maintained, but the risk is eliminated. In particular, he presents subjects with representationally identical situations in which the payment is based on the "average box", so that the boxes with different reward magnitudes contain a sure amount represented in a complex fashion. He documents virtually identical "probability distortions" across the two situations, thus showing that what has traditionally been thought of as risk attitudes or preferences may actually reflect (at least in part) attitudes towards complexity.

Garagnani and Vieider (2025) test the implications of several of the models above specifically with regards to their predictions about stochastic choice. Other than for risk-taking in general—where the predictions from different (classes of) models often diverge significantly—virtually all of the models examined predict that observed decision noise should be lower in numerical ranges to which subjects are adapted, than for numerical ranges which are encountered less frequently.Garagnani and Vieider (2025) use natural variation in currency units as a test for such adaptation. In particular, the purchasing power of 1 Great British Pound corresponds to that of about 180 Japanese Yen, and the value of 1 Euro to the value of about 400 Hungarian Florints. To ensure experimental control and to warrant causal inference, they include 2 conditions in each country: one in which subjects make decisions over the numerical ranges corresponding to typical daily purchases in their currency units, and one in which the value is maintained constant, but the experimental currency units use numerical ranges typical of the *other* country. Investigating errors such as stochastic dominance violations in risk taking, they find that subjects in the UK and Austria make significantly more mistakes for larger numerical units. In Japan and Hungary, however, the frequency of mistakes is highest in the *low* numerical units treatment. This supports the notion that decision-making errors are not driven purely by magnitude (larger is more complex), but rather by adaptation (unusual numerical ranges are more complex, regardless of whether the numbers are small or large).

In Oprea and Vieider (2024) we provide a test of the noisy cognition explanation of probability weighting and the description-experience gap that is informed by the Bayesian model described in some detail in the previous section. Arguably, however, the test goes beyond the specifics of that model, and generally illustrates how probability weighting is driven by cognitive frictions affecting the understanding of probabilities. In the base-

line condition, replicating standard setups used to test description- and experience-based choice, we replicate the finding that relative risk aversion increases in probabilities when probabilities are described, but *decreases* in probabilities when they have to be learned by sampling. We then introduce a forced sampling treatment into decisions-from-experience, obliging subjects to sample the complete urn without replacement before taking a decision. Choice patterns thereupon converge to broadly neo-classical behaviour, exhibiting mild risk aversion, but no likelihood-dependence.

Crucially, however, this treatment does not close the gap. Neither is it predicted to do so by the model: after all, standard probability distortions in the description-based paradigm are also attributed to coding noise. We thus implement a similar treatment for described choice: even while subjects are shown a full description of the choice options, they are forced to sample the whole urn without replacement. The force of this test derives from the observation that under the lens of standard models such as PT the samples provide no additional information, given that probabilities are modelled as being objectively perceived. Our noisy coding model, on the other hand, predicts that samples should contain additional information that will be added to the neurally coded signal. We indeed find behaviour to converge to mild risk aversion, without any likelihood dependence, upon forced sampling. Comparing forced sampling in experience- and description-based settings shows that the description-experience gap has closed entirely.

These tests provide fairly strong evidence that probability weighting and decision noise are driven by cognitive frictions. They do so in fairly general settings, which arguably provide support to a whole class of noisy cognition models. Below, I review some tests that are more specific to individual settings or models.

6.2 Tests of Decision-by-Sampling

DbS is perhaps the most-tested model amongst those I have described, possibly because it is also one of the first models that have been proposed (with the exception of Robson, 2001a, which however lends itself less easily to specific empirical tests). Tests have indeed been conducted for all its elements.

Decision-by-Sampling and loss aversion. Several tests have specifically targeted the emergence of loss aversion. Loss aversion in DbS is a direct result of the differential sensi-

tivity towards gains and losses in a given range, which results from the relative frequency of gains and losses of different magnitude in the environment. A natural manipulation thus consists of experimentally exposing DMs to different ranges of gains and losses.

Walasek and Stewart (2015) jointly manipulated the range of gains and losses in an experiment. Just like predicted by the model, larger losses combined with smaller gains made loss aversion disappear or even reverse, whereas it resurfaced if subjects were shown larger gains and smaller losses. A problem with these experiments is that: 1) the adaptation and choice stimuli are the same, so that the proper *consecutio temporum* between cause and effect is not given. That is, assuming that the results are driven by (perfect) adaptation requires a degree of magical thinking, since some of the choice stimuli would have been presented at the outset of the experiment, when subjects have seen few if any choices (see discussion for a more general critique of the literature arising from this); and 2) the test stimuli are not held constant across treatment conditions.

André and de Langhe (2021b) indeed show that the effects reported by Walasek and Stewart (2015) are largely due to the use of test stimuli that differ across treatments. In particular, they show 1) that the same results can be reproduced from synthetic, simulated data without treatments when tests are executed on the different test stimuli used by Walasek and Stewart (2015); and 2) that the results disappear in the original data when the same tests are executed only on stimuli that are *common* across treatments. Walasek, Mullett and Stewart (2021) nevertheless show nonparametric differences in the common stimuli, which did however not assuage their critics (see André and de Langhe, 2021a). As we will see below, the stimuli they use are not well suited to distinguish the predictions of DbS from models such as the one of Khaw et al. (2021) (which one cannot hold against them since that model did not yet exist at the time).

Bouchouicha et al. (2025) pitch DbS directly against the noisy cognition model of Khaw et al. (2021) to determine which of the two models better accounts for acceptance or rejection decisions of even odds gain-loss gambles. They use an adaptation phase to manipulate the distribution of either only gains or only losses, followed by a common test phase. The diagnostic treatment distinguishing between DbS and the noisy cognition model of Khaw et al. (2021) consists in manipulating the distribution of gains only (this is done in a pure gain setting, to avoid confounds arising when gains and losses are

manipulated jointly). Subjects shown large gains in the adaptation phase are predicted by DbS to be more risk averse than subjects shown small gains, since the size of the gains in the environment will directly impact the rank attributed to a given gain x in the common test set. The noisy cognition model, however, predicts the exact opposite: larger gains in the adaptation phase ought to shift the mean of the prior upwards, and thus reduce λ (endogenized loss aversion). The data show a clear increase in the acceptance of mean-preserving spreads around 0, consistent with the noisy cognition model, but in contradiction to DbS.

Decision-by-Sampling and probability distortions. Stewart, Reimers and Harris (2014) present systematic experimental manipulations of choice quantities, including several manipulations of the probabilities to which subjects are exposed. They then use parametric estimation methods to show that they can thereby manipulate probability weighting functions at will in a way that agrees with the predictions of DbS (they report similar results for utility as well). In a subsequent adversarial collaboration, however, Alempaki, Canic, Mullett, Skylark, Starmer, Stewart and Tufano (2019) show that the patterns described by Stewart et al. (2014) are purely an artifact of the (fairly complex) parametric analysis techniques adopted: when looking at the nonparametric data, none of the model predictions are supported. At the very least, this leaves the explanation provided by DbS in limbus in terms of its empirical validity.

A different possibility then consists in directly examining the validity of the underlying arguments. The different distribution of credits and debits, used by Stewart et al. (2006) to micro-found decreasing sensitivity towards outcomes and loss aversion, provides solid foundations on which to build adaptive models, and arguably holds significance beyond the modelling specifics adopted in the paper (see also discussion below). The data shown to support probability distortions by Stewart et al. (2006), however, are arguably not quite as convincing as their quantitative payment data for gains and losses.

Their principal method for assessing probability distributions consists in a) having some experimental subjects rate the numerical likelihoods entailed by verbal phrases (such as "most likely"; "usually"; "rarely", or "almost impossible"); and b) analyzing the frequency of occurrence of these phrases in natural language. They conclude that phrases describing very small probabilities and very large probabilities are most frequent. This argument,

however, relies crucially on the phrases chosen. While the list includes terms such as "maybe", "even odds" and "fifty-fifty chance", it does not include phrases such as "I do not know", "it is impossible to predict", or "no idea". Applied to binary events such as rain versus no rain, such phrases are, however, both likely to indicate approximate 50-50 guesses, and to be very frequent (especially so in the English climate).

Another argument they use concerns the purported higher frequency of extreme probabilities in experiments measuring probability weighting. Whether such a bias towards extreme probabilities in measurement exists is, however, not so clear. While some experiments indeed over-sample from large and small probabilities (Gonzalez and Wu, 1999, being the example typically cited), many include a multiplicity of intermediate probability tasks to produce the stake variation needed for the identification of utility curvature (L'Haridon and Vieider, 2019). A careful analysis of the universe of PT estimations as a function of probabilities deployed in the experiment seems desirable here.

6.3 Context-dependence of choices under risk

The models of Khaw et al. (2021) for utility curvature and of Vieider (2024) for probability weighting are explicitly geared towards binary choice. Applications of the same cognitive principles to certainty equivalents or valuations will generally produce different predictions. Khaw et al. (2023) and Bouchouicha et al. (2024) propose models that are tailor-made for choice lists or valuation tasks: the predictions of these models differ markedly from those of the binary choice models. Other models have been empirically applied to valuation tasks (e.g. Zhang et al., 2020; Frydman and Jin, 2023), but their more stylized nature and their application purely to the likelihood dimension means that they do not distinguish between different decision contexts or elicitation frameworks. Finally, models such as those of Netzer (2009), Herold and Netzer (2023) and Netzer et al. (2024) are formulated at a higher level of abstraction, and make no predictions on effects of the specific decision context, either.

What is increasingly clear based on the empirical literature, however, is that the way in which choices are presented can matter hugely. Bouchouicha et al. (2024) show that probability distortions obtained using choice lists (certainty equivalents) differ substantially from probability distortions in binary choice, even while keeping the underlying choices

identical across contexts. Whereas CEs produce a fourfold pattern of risk attitudes as discussed in PT, binary choice results in a *two-fold* pattern of apparent risk attitudes: risk aversion for gains, and risk seeking for losses. This is accompanied by attenuated likelihood-insensitivity in binary choice. Based on a meta-analysis they further show that the same pattern has been present in virtually all PT measurements all along, but has largely gone undetected due to a focus on structural estimations of PT functionals.

On the one hand, these findings show the limitations arising from the universalist aspirations of PT: models aiming to capture behaviour purely by applying preference functionals to objectively perceived choice primitives cannot possibly organize this sort of context-dependence. On the other hand, the results help discriminate between different models of cognitive frictions. In particular, the Bayesian Inference Framework is unique amongst the models above in having been tailored from the outset to a specific choice context. While Vieider (2024) describes probability distortions in a binary choice setup, the models of Khaw et al. (2023) and Bouchouicha et al. (2024) are explicitly geared towards probability distortions arising in valuation tasks. Although these two models have been developed in parallel and independently from each-other (as witnessed by them using a different formal angle of attack to the problem), they both make very similar predictions, and they both can account for the discrepancies between choice and valuations described in Bouchouicha et al. (2024).

This is not to say, however, that these models are the only ones that can account for these phenomena. Shubatt and Yang (2024) model differences between choice and valuation by including noise deriving from a tendency towards the center of a choice list or valuation interval. They thus organize a variety of phenomena, ranging from inversions of probability weighting when using probability equivalents instead of certainty equivalents[12], to classical preference reversals (Slovic and Lichtenstein, 1968; Lichtenstein and Slovic, 1971). Notice that these same phenomena can be organized also by the models of Khaw et al. (2023) and Bouchouicha et al. (2024), albeit by different mechanisms. Ultimately, the jury is still out on which type of cognitive friction may capture this sort of behavioural

[12]These type of reversals have been known at least since Hershey and Schoemaker (1985). The latter explained these discrepancies in the context of internal reference points in PT, with the sure amount in probability-equivalent lists acting as an endogenous reference point. It has, however, been known for some time that this does not provide a satisfactory explanation of the phenomenon—see e.g. Feldman and Ferraro (2023) for a recent examination of these issues.

regularity best, and specific tests designed to be diagnostic of the differences between the models may be required to answer this question.

6.4 Cognitive and neural correlates of decision noise

As we have seen above, all the models are motivated by some sort of cognitive friction or limitation. Most of the models furthermore appeal to some sort of neural coding as a source of (at least part of) these frictions. This opens the field for the hunt of cognitive and neural correlates of behavioural deviations from economic optimality benchmarks.

Barretto-García, de Hollander, Grueschow, Polania, Woodford and Ruff (2023) investigate the neural and number-discrimination correlates of outcome distortions as modelled by Khaw et al. (2021). They quantify neural signatures of the precision of number representation in parietal cortex, and document correlations between the neural accuracy indices and numerical discrimination tasks using either clouds-of-dots displays to represent magnitudes or symbolic representations (Arabic numerals). Since according to the model of Khaw et al. (2021) such numerical discrimination lies at the heart of risky choice, they also investigate correlations with risky choice. The neural measures of coding precision are indeed found to correlate with performance in numerical discrimination tasks. They also correlate with behavioural risk aversion, both in tasks using non-symbolic and in tasks using symbolic representations.

Neural correlates have also been studied for attitudes towards mixed gambles. In a seminal study, Tom et al. (2007) measured neural activation functions while subjects made accept-reject decisions of binary gain-loss wagers. They showed that deactivations measured over a range of losses were stronger than activations over a range of gains. The differential reactions to gains versus losses were furthermore highly predictive of acceptance decisions of 50-50 gain-loss gambles. This provides neural evidence for the concept of loss-sensitivity enshrined in the model of Khaw et al. (2021).

The neural results of Tom et al. (2007) suggest that such loss-sensitivity ought to be driven by increased attention towards losses relative to gains. Pachur et al. (2018) directly test such an attentional account using a Mouselab paradigm, where subjects have to hover over different attributes with the mouse to be able to see them. They thereby conceive of attention as the time spent on a given attribute. Time spent considering losses relative to

gains is indeed found to be highly predictive of loss aversion parameters estimated in a PT model. They also exogenously vary attention by showing gains versus losses for different lengths of time. Although the effects go in the expected direction, thus providing some causal evidence for the effect of attention, the recoded effects are on the weak side from both a statistical and a substantive point of view (measured loss aversion hovers around 1 in all cases). Hirmas et al. (2024) further investigate the same issue using rich eye-tracking data. Bouchouicha et al. (2025) show that a stylized attention model injected into KLW can capture such attentional effects, whereby the asymmetry in attention to gains and losses will drive loss-sensitivity.

Bouchouicha et al. (2025) further correlate relative sensitivity towards losses and gains with measures of cognitive and numerical acuity. Loss-sensitivity—capturing the difference in attention to losses and gains—is shown to decrease in cognitive ability. They trace this effect to increased attention towards gains by cognitively more able people, thereby explaining the opposite correlations of cognitive ability with risk aversion over pure gains and mixed gambles reported by Chapman, Snowberg, Wang and Camerer (2024). Similar correlations of cognitive ability with likelihood-sensitivity have been reported by a number of papers (see e.g. L'Haridon and Vieider, 2019; Choi, Kim, Lee, Lee et al., 2022). All of this is indicative that cognitive ability may contribute to determining the precision of noisy signals overall. Enke and Graeber (2023) document behavioural attenuation that is common across risk taking and beliefs, as well as forecasts (see also Zhang and Maloney, 2012). Behavioural attenuation under risk—in the sense of behaviour that seems driven by probabilities that lie closer to an intermediate mean than the objective probabilities suggest, thus resulting in probability weighting—is correlated with answers given to a simple survey question asking subjects how certain they are of their choice. This suggests that subjects have some conscious awareness of their cognitive noise.

7 Current limitations and future opportunities

From the extensive if somewhat selective review I have presented in this chapter, it is clear that all the models agree on some fundamental aspects driving decisions. The first element consists in cognitive limitations that lead to frictions in the decision-making process. The second element consists in some way of dealing with the cognitive frictions

arising in the first step. Usually, this step is represented as being "optimal" . The models do, however, differ in where the frictions are supposed to arise precisely, how they affect decisions, and how precisely such frictions are dealt with. They may also differ in terms of their optimality criteria. It is not always clear by looking at the models where the differences arise, and which differences may actually matter in terms of the predictions they make. Here, I first present a discussion of commonalities and differences, and how they may affect predicted behaviour. Subsequently, I present a more general discussion of the current limitations of this literature, and of its future promise.

7.1 Models of encoding and models of decoding

It is difficult to find a single dimension along which the models can be classified. An important dimension nevertheless seems to be the distinction between encoding and decoding. Encoding describes the process whereby the choice primitives in the real world are neurally represented. The decoding subsequently consists in deciphering the information content of the codes, and using them towards forming a decision. Note that one could model additional stages (e.g. distinguishing purely perceptual encoding from recombination of choice quantities, all of which may produce noise), and that some models may not fit this scheme well (possibly because their justification is not explicitly neural). Many models may further include both stages, but differ with regards to where "the main action" occurs when it comes to predicting behaviour. This simple dichotomy will nevertheless allow us to start a discussion about the commonalities and differences between the models discussed above.

A typical example of a model of encoding is provided by Thurstone (1927a), who models discrimination between two stimuli based directly on the signals representing them. Such signals will typically be noisy, given that under plausible modelling assumptions one can show that an infinity of neurons would be needed for perfectly accurate presentations. The divisive normalization model of Glimcher and Tymula (2023) is an example where predictions are explicitly based on adaptation of neural firing rates to the stimuli in the environment. The action therefore derives from optimal noise adaptation, subject to finite information-processng resources. More recent iterations of efficient coding models, such as the probability weighting explanation proposed by Frydman and Jin (2023), share the feature that the main action derives from the neural encoding stage.

The encoding models discussed above share the common feature that the main action derives from the neural or sensory apparatus adjusting to reflect the distribution of stimuli in the environment. This is a feature that they share with the evolutionary setup of Robson-Netzer, and with the Decision-by-Sampling framework. A special feature these models have is that outcomes and probabilities are assumed to be perceived perfectly and without noise. This feature may seem logically at odds with the very motivation of neural encoding models (although it is a feature that is shared by the explicitly neural model of Glimcher and Tymula, 2023). Noise nevertheless arises in the attribution of decision values, such as utilities to outcomes, to such objectively perceived choice primitives. The observation that the utility-attribution process is once again driven by efficiency concerns similar to those detailed in the encoding models above justifies classifying these two model groups alongside each other as models of "efficient coding".

Decoding models specifically put the action in the decoding process. Encoding is thereby typically seen as noisy, but noise is often—although not necessarily—treated as exogenous to the model. This class of models includes notably Bayesian inference (or observer) models such as the Khaw et al. (2021), Khaw et al. (2023), and Vieider (2024). Here, coding noise is often treated as uniform over the stimulus space (although this is mostly a simplifying feature in these models; see e.g. Zhang et al., 2020, for an exception). While the main action thus derives from the decoding stage, adding optimal adaptation in encoding to these models can yield additional predictions.

Some models derive their predictions from the interaction of the encoding and decoding stages. The sampling-based model of Oprea and Vieider (2024)—even though it builds on and generalizes the model of Vieider (2024)—crucially relies on the endogenization of coding noise in experience-based choices. It then is the combination of coding noise and regression to the Bayesian prior mean which drives the endogenous sampling-stopping rule in the model. The stopping decision, in turn, predicts behaviour. Netzer et al. (2024) explicitly characterize both the encoding and decoding stages in valuations of lotteries with a multiplicity of states, and show how noise in encoding and decoding can yield separate behavioural predictions.

7.2 Models of adaptation

Considering that almost all models I have reviewed rely on an adaptive mechanism to counter-balance cognitive bottlenecks, the explicit modelling of adaptation is conspicuously absent from most models (see, however, Glimcher and Tymula, 2023, for a model that is explicitly adaptive; see also Robson, Whitehead and Robalino, 2023, for a model of adaptation specific to the Robson-Netzer framework). For instance, empirical tests of DbS have relied on manipulations of the choice statistics in the environment, but the extent to which distributions in the immediate experimental environment may add to or even over-write distributions learned in the real world remains unclear due to the absence of an explicit model of memory formation. Similar issues occur for the Bayesian observer models, where learning of the prior has received relatively little attention.

The problems arising from the neglect of learning are bigger than one might at first think. Given the noisy perception of outcomes, assuming that the environmental stimuli are learned perfectly will simply not cut it. To the extent that choice primitives (or their values) are themselves noisily perceived, a perfect learning assumption risks to result in magical thinking. In particular, any systematic bias in perception will likely be reflected in learned distributions, although the extent to which this happens, and the degree of distortion, will depend on the fine details of the process. This is particularly problematic in models such as DbS and efficient coding, which often fundamentally rely on the assumption of correct learning of the stimuli in the environment to derive their predictions of noise adaptation. If learning is not only noisy but systematically biased—as some of the empirical results in the literature do indeed suggest—then the failure of this assumption threatens the very core of those models. Building explicit models of such learning processes should thus be a priority task for future research.

7.3 Cognitive frictions and the status of preferences

Some of the models and empirical results I have presented in this chapter have been interpreted to imply that stable preferences do not exist. This impression can easily arise from the focus put on effects arising from complexity, as in **?**, or from the optimal choice rules which constitute the starting point of Khaw et al. (2021) and Vieider (2024), assuming expected value maximization. This impression, however, is incorrect. The point

of both the empirical investigations and the models is more accurately that any stable preferences may not be easily extracted from observed choices, given that they will be confounded by mistaken inferences arising from noisy cognitive processes.

An interesting question is, nevertheless, what form preferences might take in this framework (and as a consequence, how one could measure them). One possibility is simply to allow for a standard utility function applied to the objective choice primitives, such as explicitly discussed by both Khaw et al. (2021) and Vieider (2024). True preferences would then take the form of decreasing marginal utility of wealth, but would likely not impact small-stake decisions, which arise purely from noisy cognition (the key point on Khaw et al. (2021)). In principle, however, even stable, preference-driven utility over small stakes, such as proposed by Alaoui and Penta (2025), could be integrated into models of noisy cognition without loss of generality.

Another intriguing possibility is to look for relatively stable components of choice within the generative frameworks themselves. For instance the mean of the Bayesian prior could be learned very conservatively in Bayesian Inference Models, reflecting a hierarchical structure (Friston, 2005) that is learned over a lifetime and changes only slowly based on noisy inferences. The same holds for environmental distributions of choice stimuli such as modelled in DbS, where experiences in certain phases of life (e.g. formative years) could have a disproportionate influence. The great advantage of such an endogenous account is that it could potentially offer a way to study preference *formation*—something the accounts discussed above cannot. That being said, a meaningful discussion of these problems requires the development of formal models of learning of environmental distributions, which are at present still lacking from the literature.

7.4 Conclusion: Unified foundations of decision-making

Prospect theory has been an extremely successful theory of choice. Over time, however, some limitations have emerged. I argued that these limitations arise 1) from the universalist aspirations of prospect theory, which prevent it from accounting for the a variety of procedure invariance violations that have been documented in the empirical literature; and 2) from the *ex post* nature of the fitting exercise, which results in a large (and in extreme cases: *infinite*) number of *ex post* parameters, which prevent it from making

meaningful predictions. In this chapter, I have thus argued for the promise of cognitive models that endogenize PT-like parameters to overcome both these issues.

The promise of models in the cognitive tradition goes beyond the narrow focus of decisions under risk I reviewed here. Models such as Bayesian Inference and Efficient Coding rest on fundamental principles that are thought to underly neural processes in general, thus unifying the modelling of higher cognitive functions with that of sensori-motor tasks. Very similar—or at times even identical—models to those described here can furthermore be used to account for decision-making patterns under certainty, for ambiguity attitudes, delay-discounting, and possibly even for social interactions. This drive for unification—jointly with the relative parsimony of the approach—is one of the greatest promises of the neuro-cognitive approach to decision modelling.

References

Abdellaoui, Mohammed, Aurélien Baillon, Lætitia Placido, and Peter P. Wakker (2011) 'The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation.' *American Economic Review* 101, 695–723

Alaoui, Larbi, and Antonio Penta (2025) 'What's in a u?' Technical Report, Universitat Pompeu Fabra

Alempaki, Despoina, Emina Canic, Timothy L Mullett, William J Skylark, Chris Starmer, Neil Stewart, and Fabio Tufano (2019) 'Reexamining how utility and weighting functions get their shapes: A quasi-adversarial collaboration providing a new interpretation.' *Management Science* 65(10), 4841–4862

André, Quentin, and Bart de Langhe (2021a) 'How (not) to test theory with data: Illustrations from walasek, mullett, and stewart (2020).'

_ (2021b) 'No evidence for loss aversion disappearance and reversal in walasek and stewart (2015).' *Journal of Experimental Psychology: General* 150(12), 2659

Apesteguia, Jose, and Miguel A Ballester (2018) 'Monotone stochastic choice models: The case of risk and time preferences.' *Journal of Political Economy* 126(1), 74–106

Atchison, J, and Sheng M Shen (1980) 'Logistic-normal distributions: Some properties and uses.' *Biometrika* 67(2), 261–272

Barretto-García, Miguel, Gilles de Hollander, Marcus Grueschow, Rafael Polania, Michael

Woodford, and Christian C Ruff (2023) 'Individual risk attitudes arise from noise in neurocognitive magnitude representations.' *Nature Human Behaviour* 7(9), 1551–1567

Birnbaum, Michael H (1999) 'Testing critical properties of decision making on the internet.' *Psychological Science* 10(5), 399–407

_ (2008) 'New paradoxes of risky decision making.' *Psychological review* 115(2), 463

Bishop, Christopher M. (2006) *Pattern recognition and machine learning,* vol. 4 (Springer)

Bossaerts, Peter, Nitin Yadav, and Carsten Murawski (2019) 'Uncertainty and computational complexity.' *Philosophical Transactions of the Royal Society B* 374(1766), 20180138

Bouchouicha, Ranoua, and Ferdinand M. Vieider (2017) 'Accommodating stake effects under prospect theory.' *Journal of Risk and Uncertainty* 55(1), 1–28

Bouchouicha, Ranoua, Ryan Oprea, Ferdinand M. Vieider, and Jilong Wu (2024) 'Is prospect theory really a theory of choice?' Technical Report, Ghent University Discussion Papers

Bouchouicha, Ranoua, Yuchi Li, and Ferdinand M. Vieider (2025) 'Loss-sensitivity versus loss-aversion.' Technical Report, Ghent University Discussion Papers

Chapman, Jonathan, Erik Snowberg, Stephanie W Wang, and Colin Camerer (2024) 'Looming large or seeming small? attitudes towards losses in a representative sample.' *Review of Economic Studies, forthcoming*

Choi, Syngjoo, Jeongbin Kim, Eungik Lee, Jungmin Lee et al. (2022) 'Probability weighting and cognitive ability.' *Management Science, forthcoming* 68(7), 4755–5555

Dehaene, Stanislas (2003) 'The neural basis of the weber–fechner law: a logarithmic mental number line.' *Trends in cognitive sciences* 7(4), 145–147

_ (2011) *The number sense: How the mind creates mathematics* (OUP USA)

Drugowitsch, Jan, Valentin Wyart, Anne-Dominique Devauchelle, and Etienne Koechlin (2016) 'Computational precision of mental inference as critical source of human choice suboptimality.' *Neuron* 92(6), 1398–1411

Ellsberg, Daniel (1961) 'Risk, Ambiguity and the Savage Axioms.' *Quarterly Journal of Economics* 75(4), 643–669

Enke, Benjamin, and Cassidy Shubatt (2023) 'Quantifying lottery choice complexity.' Technical Report, Mimeo

Enke, Benjamin, and Thomas Graeber (2023) 'Cognitive uncertainty.' *Quarterly Journal*

of Economics

Enke, Benjamin, Thomas Graeber, Ryan Oprea, and Jeffrey Yang (2024) 'Behavioural attenuation.' Mimeo

Ert, Eyal, and Ido Erev (2013) 'On the descriptive value of loss aversion in decisions under risk: Six clarifications.' *Judgment and Decision making* 8(3), 214–235

Fechner, Gustav Theodor (1860) (Kessinger's Legacy Reprints)

Fehr-Duda, Helga, Adrian Bruhin, Thomas F. Epper, and Renate Schubert (2010) 'Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size.' *Journal of Risk and Uncertainty* 40(2), 147–180

Feldman, Paul J, and Paul J Ferraro (2023) 'A certainty effect for preference reversals under risk: Experiment and theory'

Friedman, Daniel (1989) 'The s-shaped value function as a constrained optimum.' *The American Economic Review* 79(5), 1243–1248

Friedman, Milton, and L. J. Savage (1948) 'The Utility Analysis of Choices Involving Risk.' *Journal of Political Economy* 56(4), 279–304

Friston, Karl (2005) 'A theory of cortical responses.' *Philosophical transactions of the Royal Society B: Biological sciences* 360(1456), 815–836

Frydman, Cary, and Lawrence J Jin (2022) 'Efficient coding and risky choice.' *Quarterly Journal of Economics* 136, 161–213

_ (2023) 'On the source and instability of probability weighting.' Technical Report, National Bureau of Economic Research

Garagnani, Michele, and Ferdinand M. Vieider (2025) 'Economic consequences of numerical adaptation.' *Psychological Science, forthcoming*

Ghirardato, P, and M Marinacci (2002) 'Ambiguity made precise: A comparative foundation.' *Journal of Economic Theory* 102, 251–289

Gilboa, Itzhak (1987) 'Expected utility with purely subjective non-additive probabilities.' *Journal of Mathematical Economics* 16(1), 65–88

Glimcher, Paul W, and Agnieszka A Tymula (2023) 'Expected subjective value theory (esvt): A representation of decision under risk and certainty.' *Journal of Economic Behavior & Organization* 207, 110–128

Gold, Joshua I, and Michael N Shadlen (2001) 'Neural computations that underlie decisions about sensory stimuli.' *Trends in cognitive sciences* 5(1), 10–16

_ (2002) 'Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward.' *Neuron* 36(2), 299–308

Goldstein, W, and H Einhorn (1987) 'Expression Theory and the Preference Reversal Phenomena.' *Psychological Review* 94, 236–254

Gonzalez, Richard, and George Wu (1999) 'On the Shape of the Probability Weighting Function.' *Cognitive Psychology* 38, 129–166

Green, David Marvin, John A Swets et al. (1966) *Signal detection theory and psychophysics,* vol. 1 (Wiley New York)

Heng, Joseph A, Michael Woodford, and Rafael Polania (2020) 'Efficient sampling and noisy decisions.' *Elife* 9, e54962

Herold, Florian, and Nick Netzer (2023) 'Second-best probability weighting.' *Games and Economic Behavior* 138, 112–125

Hershey, John C., and Paul J. H. Schoemaker (1985) 'Probability versus Certainty Equivalence Methods in Utility Measurement: Are They Equivalent?' *Management Science* 31(10), 1213–1231

Hertwig, Ralph, Greg Barron, Elke U Weber, and Ido Erev (2004) 'Decisions from experience and the effect of rare events in risky choice.' *Psychological science* 15(8), 534–539

Hirmas, Alejandro, Jan B Engelmann, and Joël van der Weele (2024) 'Individual and contextual effects of attention in risky choice.' *Experimental Economics* pp. 1–28

Holt, Charles A., and Susan K. Laury (2002) 'Risk Aversion and Incentive Effects.' *American Economic Review* 92(5), 1644–1655

Imai, Taisuke, Salvatore Nunnari, Jiling Wu, and Ferdinand M. Vieider (2025) 'Meta-analysis of prospect theory parameters.' *Working Paper*

Kahneman, Daniel, and Amos Tversky (1979) 'Prospect Theory: An Analysis of Decision under Risk.' *Econometrica* 47(2), 263 – 291

Khaw, Mel Win, Ziang Li, and Michael Woodford (2021) 'Cognitive imprecision and small-stakes risk aversion.' *The Review of Economic Studies* 88(4), 1979–2013

_ (2023) 'Cognitive imprecision and stake-dependent risk attitudes.' Technical Report

L'Haridon, Olivier, and Ferdinand M. Vieider (2019) 'All over the map: A worldwide comparison of risk preferences.' *Quantitative Economics* 10, 185–215

L'Haridon, Olivier, Ryan Oprea, Rafael Polania, and Ferdinand M. Vieider (2023) 'Cognitive foundations of ambiguity atttitudes.' *Mimeo*

Lichtenstein, Sarah, and Paul Slovic (1971) 'Reversals of preference between bids and choices in gambling decisions.' *Journal of experimental psychology* 89(1), 46

Markowitz, Harry (1952) 'The Utility of Wealth.' *Journal of Political Economy* 60(2), 151–158

Mosteller, Frederick, and Philip Nogee (1951) 'An experimental measurement of utility.' *Journal of political economy* 59(5), 371–404

Natenzon, Paulo (2019) 'Random choice and learning.' *Journal of Political Economy* 127(1), 419–457

Netzer, Nick (2009) 'Evolution of time preferences and attitudes toward risk.' *American Economic Review* 99(3), 937–55

Netzer, Nick, Arthur Robson, Jakub Steiner, and Pavel Kocourek (2024) 'Risk perception: measurement and aggregation.' *Journal of the European Economic Association* p. jvae053

Oprea, Ryan (2024) 'Decisions under risk are decisions under complexity.' *American Economic Review*

Oprea, Ryan, and Ferdinand M. Vieider (2024) 'Minding the gap: On the origins of the description-experience gap.' *Working Paper*

Pachur, Thorsten, Michael Schulte-Mecklenbeck, Ryan O Murphy, and Ralph Hertwig (2018) 'Prospect theory reflects selective allocation of attention.' *Journal of experimental psychology: general* 147(2), 147

Raiffa, Howard (1961) 'Risk, Ambiguity and the Savage Axioms: Comment.' *Quarterly Journal of Economics* 75(4), 690–694

Robson, Arthur J (2001a) 'The biological basis of economic behavior.' *Journal of Economic Literature* 39(1), 11–33

_ (2001b) 'Why would nature give individuals utility functions?' *Journal of Political Economy* 109(4), 900–914

Robson, Arthur J, Lorne A Whitehead, and Nikolaus Robalino (2023) 'Adaptive utility.' *Journal of Economic Behavior & Organization* 211, 60–81

Rottenstreich, Yuval, and Christopher K. Hsee (2001) 'Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk.' *Psychological Science* 12(3), 185–190

Shubatt, Cassidy, and Jeffrey Yang (2024) 'Similarity and comparison complexity.' *arXiv preprint arXiv:2401.17578*

Slovic, Paul, and Sarah Lichtenstein (1968) 'Relative importance of probabilities and payoffs in risk taking.' *Journal of experimental psychology* 78(3p2), 1

Steiner, Jakub, and Colin Stewart (2016) 'Perceiving prospects properly.' *American Economic Review* 106(7), 1601–31

Stewart, Neil, Nick Chater, and Gordon DA Brown (2006) 'Decision by sampling.' *Cognitive psychology* 53(1), 1–26

Stewart, Neil, Stian Reimers, and Adam JL Harris (2014) 'On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes.' *Management Science* 61(3), 687–705

Thurstone, Louis L (1927a) 'A law of comparative judgment.' *Psychological review* 34(4), 273

_ (1927b) 'Psychophysical analysis.' *The American Journal of Psychology* 38(3), 368–389

Tom, Sabrina M., Craig R. Fox, Christopher Trepel, and Russell A. Poldrack (2007) 'The Neural Basis of Loss Aversion in Decision-Making Under Risk.' *Science* 315(5811), 515–518

Tversky, Amos (1969) 'Intransitivity of preferences.' *Psychological review* 76(1), 31

Vickrey, William (1945) 'Measuring Marginal Utility by Reactions to Risk.' *Econometrica* 13(4), 319

Vieider, Ferdinand M. (2024) 'Decisions under uncertainty as bayesian inference on choice options.' *Management Science* pp. 9014–9030

Walasek, Lukasz, and Neil Stewart (2015) 'How to make loss aversion disappear and reverse: tests of the decision by sampling origin of loss aversion.' *Journal of experimental psychology: general* 144(1), 7

Walasek, Lukasz, Timothy L Mullett, and Neil Stewart (2021) 'Acceptance of mixed gambles is sensitive to the range of gains and losses experienced, and estimates of lambda (λ) are not a reliable measure of loss aversion: Reply to andré and de langhe (2021).'

Wilcox, Nathaniel T. (2011) ''Stochastically more risk averse:' A contextual theory of stochastic discrete choice under risk.' *Journal of Econometrics* 162(1), 89–104

Woodford, Michael (2012) 'Prospect theory as efficient perceptual distortion.' *American Economic Review* 102(3), 41–46

Wu, George, and Alex B. Markle (2008) 'An Empirical Test of Gain-Loss Separability in

Prospect Theory.' *Management Science* 54(7), 1322–1335

Wulff, Dirk U, Max Mergenthaler-Canseco, and Ralph Hertwig (2018) 'A meta-analytic review of two modes of learning and the description-experience gap.' *Psychological bulletin* 144(2), 140

Zhang, Hang, and Laurence T Maloney (2012) 'Ubiquitous log odds: a common representation of probability and frequency distortion in perception, action, and cognition.' *Frontiers in neuroscience* 6, 1

Zhang, Hang, Xiangjuan Ren, and Laurence T Maloney (2020) 'The bounded rationality of probability distortion.' *Proceedings of the National Academy of Sciences* 117(36), 22024–22034

Zheng, Jieyu, and Markus Meister (2024) 'The unbearable slowness of being: Why do we live at 10 bits/s?' *Neuron*