What are we weighting for?

A mechanistic model for probability weighting *

Ole Peters† Alexander Adamou‡ Mark Kirstein§
Yonatan Berman¶

December 11, 2020

Abstract

Probability weighting expresses a mismatch between probabilities specified in a model and weights inferred from human decisions. Typically weights are larger than probabilities for rare events and smaller for common events in experiments. This is presented as a cognitive bias of decision makers. We provide an alternative interpretation: decision makers and observers have different uncertainty about events. We offer a mechanism generating differences in uncertainty: when a decision maker estimates probabilities as frequencies while the observer knows them *a priori*. We argue that many phenomena labelled as probability weighting may be principled responses to uncertainties unaccounted for in observer models. (*JEL* C61, D01, D81)

*The code to generate the figures is available at https://bit.ly/lml-pw-code. We benefited from comments by seminar participants at the Max Planck Institute (MPI) for Human Development, Berlin, MPI for Mathematics in the Sciences, Leipzig, MPI for Evolutionary Anthropology, Leipzig, Institute of Mathematics, Leipzig University, Erich-Schneider-Research-Seminar at Christian-Albrecht University of Kiel, Institute for Socioeconomics at University of Duisburg-Essen, and from conference participants at the first scientific workshop of the German Network for Pluralism in Economics and the D-TEA Conference 2020.

†London Mathematical Laboratory and Santa Fe Institute, ✉ o.peters@lml.org.uk

‡London Mathematical Laboratory, ✉ a.adamou@lml.org.uk

§London Mathematical Laboratory, Institute of Mathematics at Leipzig University and Max-Planck-Institute for Mathematics in the Sciences, Leipzig, ✉ m.kirstein@lml.org.uk

¶London Mathematical Laboratory and Stone Center on Socio-Economic Inquality at City University New York, ✉ y.berman@lml.org.uk

1 Introduction

"Probability weighting" is a concept that originated in psychology, going back at least to Preston and Baratta (1948). It was later popularised as a key feature of models in behavioural economics, such as prospect theory (Kahneman and Tversky 1979) and cumulative prospect theory (Tversky and Kahneman 1992). Consider a thought experiment, in which

- a *disinterested observer* (DO, he/him), such as an experimenter, tells
- a *decision maker* (DM, she/her)

that an event occurs with some probability. The DO observes the DM's behaviour (*e.g.*, gambling on the event) and finds it consistent with a behavioural model (*e.g.*, expected utility theory) in which she uses a probability that differs systematically from what he declared. The apparent probabilities, inferred from the DM's decisions, are called "decision weights," and the apparent conversion by the DM of declared probabilities to decision weights is labelled probability weighting.

We adopt a streamlined nomenclature here. Let x be a realisation of the random variable, X. For example, X might be the payout of a gamble which the DM is invited to accept or decline. There are two relevant distributions of X:

- that specified by the DO, which we call the *probability* distribution; and
- that inferred by the DO from the DM's behaviour, called the *weight* distribution.

We express a probability distribution using a probability density function, $p(x)$, and a weight distribution using a weight density function, $w(x)$. Probabilities and weights for specific events are obtained as integrals of the corresponding density functions over the relevant range of x.[1]

[1]Weights are not always normalised in the literature, whereas we work with a normalised weight density. Mathematically, then, $w(x)$ is a proper probability density function, although for clarity we avoid referring to it as such. Our results are unaffected because normalising just means dividing by a constant (the integral of the un-normalised weight density).

We take as our phenomenological definition of probability weighting the robust observation that weights are higher than probabilities for rare events (and, by normalisation, lower than probabilities for common events). Specifically, in this paper we consider

- cumulative probabilities,

$$F_p(x) = \int_{-\infty}^{x} p(s)ds \ , \tag{1}$$

- and cumulative weights,

$$F_w(x) = \int_{-\infty}^{x} w(s)ds \ . \tag{2}$$

$F_p(x)$ and $F_w(x)$ are the cumulative density functions corresponding, respectively, to the probability and weight distributions of X.

In other words, we are confining our attention to probabilities and weights of events of the type $X < x$, *e.g.*, winning less than \$10 in a gamble. This is consistent with the form of probability weighting used in more recent behavioural models, such as in Tversky and Kahneman (1992). In earlier models, and in the original observations of Preston and Baratta (1948), probabilities and weights can refer to other types of events.

Figure 1 summarises the data in Preston and Baratta (1948, p. 188) (for non-cumulative probabilities and weights). In our framework, we would plot F_w as a function of F_p to generate a curve, whose most commonly observed shape we call the "inverse-S." This curve sits above the diagonal, $F_w = F_p$, for events of low cumulative probability (such that $F_w > F_p$ for these events) and below the diagonal for events of high cumulative probability (such that $F_w < F_p$).

As a final piece of nomenclature, we use the terms *location*, *scale*, and *shape* when discussing probability density functions. In a plot of a density function, a change of location shifts the curve to the left or right, and a change in scale shrinks or enlarges the width of its features. Neither operation changes the

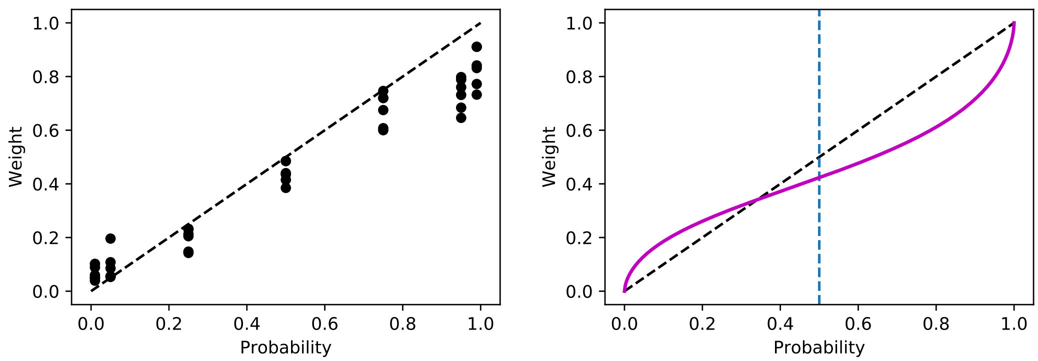

Figure 1: **Empirical phenomenon of probability weighting.** Left: weights *vs.* probabilities as reported by PRESTON and BARATTA (1948), the first to document probability weighting empirically (they refer to probabilities as "mathematical probabilities," weights as "psychological probabilities," and consider events of non-cumulative type). Low probabilities are exceeded by their corresponding weights and *v.v.* for high probabilities. Right: fitting a curve to these data produces an inverse-S, see Sec. 2.1 for details.

shape of the density function. For a general random variable X, with arbitrary location parameter μ_X and scale parameter σ_X, the following transformation produces a standardised random variable with an identically-shaped density function, but with location $\mu_Z = 0$ and scale $\sigma_Z = 1$:

$$Z = \frac{X - \overbrace{\mu_X}^{\text{location}}}{\underbrace{\sigma_X}_{\text{scale}}} \,. \tag{3}$$

Two density functions have the same shape if they can be made to coincide through a linear transformation of the form (Eq. 3).

1.1 Related literature

Formal probability theory has its roots in the 1650s as a tool to analyse gambles. These early developments led to expected value theory (EVT), a behavioural model in which people choose among gambles by optimising the expectation value of the payoff.

While successful in some circumstances, EVT led to paradoxes and puzzles. The first proposed remedy to these paradoxes was expected utility theory (EUT) (Bernoulli 1738; Laplace 1814). EUT distorts monetary amounts to reflect people's apparent psychological risk aversion, leaving the expectation-value structure of the optimand (a weighted sum) intact. Axiomatisations of EUT by von Neumann and Morgenstern (1944) and Savage (1954) helped to establish EUT as a normative decision theory.

Despite its flexibility, EUT failed to fit some empirical observations, which prompted Preston and Baratta (1948) to introduce a psychological distortion not of wealth but of probabilities. Today, this alternative distortion is known as probability weighting and is the focus of our study. Probability weighting later found its way into descriptive decision theories such as prospect theory (Kahneman and Tversky 1979) and then cumulative prospect theory (CPT) (Tversky and Kahneman 1992). Under the consensus interpretation, probability weighting is a systematic and maladaptive bias, in which people give excessive weight to rare events and insufficient weight to common events in their decisions.

As research continued, critiques of probability weighting emerged. With both monetary amounts and probabilities now transformed by freely chosen nonlinear functions, there is a risk of overfitting: parameters which fit past data well may fit new data poorly. Indeed, Stewart et al. (2015) fail to find stable mappings from objective probabilities to subjective weights in individual DMs. They find that specific mappings depend on the choice environment and are influenced to a large extent by the sampling method.

A key confounder turned out to be experimental design. Research under the heading of the *description-experience gap* revealed that it matters how DMs are informed of probabilities in a decision problem. In *decisions-from-description*, DMs are told the probabilities by DOs. In *decisions-from-experience*, DMs must infer the probabilities from their experience over time.[2]

[2]Roughly since the year 2000, experience-based experiments have prevailed (Hertwig et al. 2004; Hertwig and Erev 2009; Erev et al. 2010). In animal studies, only the

Since there are different possible physical interpretations of probabilities, see Sec. 3, decisions-from-description lack potentially relevant semantic information. Decisions-from-experience are clearer in this respect but confront the DM with the problem of induction. Hertwig et al. (2004) find different (in their case, stronger) overweighting of rare events in decisions-from-description than in decisions-from-experience. Some experience-based experiments have even found underweighting of rare events (Ungemach et al. 2009; Wulff et al. 2018, esp. Tab. 9).

Decisions-from-experience studies emphasise statistical over psychological explanations of probability weighting. Martins (2006) interprets probability weighting as a heuristic that approximates a Bayesian solution to the problem of inferring probabilities from data. In line with our analysis, Seo et al. (2019) find that optimal weights for a DM deviate from true probabilities in a sequential learning problem, such that it is optimal to overstate small probabilities and understate large ones. Fox and Hadar (2006) and Ungemach et al. (2009) report that probability weighting persists even after controlling for such effects, suggesting that a (much smaller residual) psychological component may still exist.

In short, recent studies raise questions about the reproducibility of probability weighting, its dependence on the details and context of decisions, and its interpretation as either a maladaptive psychological bias or a beneficial statistical heuristic. Our impression is that an empirical effect of apparent probability weighting exists, but it is less clear than early studies suggested, and, as we shall see, its interpretation as a maladaptive cognitive bias is questionable.

In our own work, ergodicity economics (EE), the DM's experience along a single timeline is fundamental (Peters 2019). Therefore, in this paper we begin by asking what a DM can reasonably infer about probabilities from available information, *e.g.*, a statement by a DO with no corroborating evidence, or a finite sample of observed outcomes. In decision theory as it pertains to individuals, EE rejects EVT because there is no *a priori* physical reason for

decisions-from-experience paradigm is feasible.

expectation values to reflect what happens to an individual over time. Instead, EE uses as its null model the optimisation of appropriate growth rates evaluated in the long-time limit (Peters 2011a,b, 2019). For clarity, this is in contrast to decision theories which rely on expectation values (suitably distorted) as their optimands. EE finds that growth-optimal behaviour, defined properly along a single timeline, predicts correctly a number of deviations from standard EUT (Meder et al. 2019; Adamou et al. 2020; Berman and Kirstein 2020). The mechanism presented here explains probability weighting as arising from the distinction between the models of uncertainty used by the DO (who defines the setup) and by the DM (who must infer it).

We offer a statistical explanation of probability weighting that applies to any situation where DO and DM have different uncertainties.

2 Probability weighting as a difference between models

Behavioural economics interprets Fig. 1 as evidence for a cognitive bias of the DM, an error of judgement. We will keep a neutral stance. We don't assume the DO to know "the truth" – he has a formal model. Nor do we assume the DM to know "the truth" – she has another model, perhaps less formal, on which she bases her actions. From our perspective Fig. 1 merely shows that the two models differ. It says nothing about who is right or wrong.

2.1 The inverse-S curve

Tversky and Kahneman

Tversky and Kahneman (1992) chose to fit empirical data, qualitatively similar to the data presented in Fig. 1 from Preston and Baratta (1948), with the following function,

$$F_w^{TK}(F_p;\gamma) = (F_p)^\gamma \frac{1}{\left[(F_p)^\gamma + (1-F_p)^\gamma\right]^{1/\gamma}} , \tag{4}$$

which maps from one cumulative density function (CDF), F_p, to another, F_w. We note that no mechanistic motivation was given for this specific family of CDF mappings, parametrised by γ. The motivation seems purely phenomenological: with $\gamma < 1$, this function can be made to fit to the data reasonably well.

The function $F_w^{TK}(F_p; \gamma)$ has one free parameter, γ. For $\gamma = 1$ it is the identity, and the CDFs coincide, $F_w^{TK}(F_p) = F_p$. Further, the function has the following property: any curvature moves the intersection with the diagonal away from the mid-point, $(\frac{1}{2}, \frac{1}{2})$. This means if the function is used to fit an inverse-S (where $\gamma < 1$), the fitting procedure itself introduces a shift of the intersection to the left. This is demonstrated in the right panel of Fig. 1. We consider the key observation of probability weighting to be the inverse-S shape, whereas the shift to the left may be an artefact of the function chosen for the fit.

Scale, location, and the inverse-S

We now make explicit how the robust qualitative observation of the inverse-S shape emerges when the DM uses a larger scale in her model than the DO.

We illustrate this with a Gaussian distribution. Let's assume that a DO models a future change in wealth, X, as a Gaussian random variable with location μ and variance σ^2. And let's further assume that a DM models this change as a Gaussian random variable with the same location, μ, but with a greater scale, such that the variance is $(\alpha\sigma)^2$ with $\alpha > 1$. The DM simply assumes a broader range – α times greater – of plausible values, left panel of Fig. 2.

If the DM uses a greater scale in her model, then weight densities are higher than probability densities where the densities are low, and (because of normalisation) lower than probability densities where the densities are high. We can express this by plotting, for any value of x, the weight *vs.* probability densities observed at x, as shown in the right panel of Fig. 2.

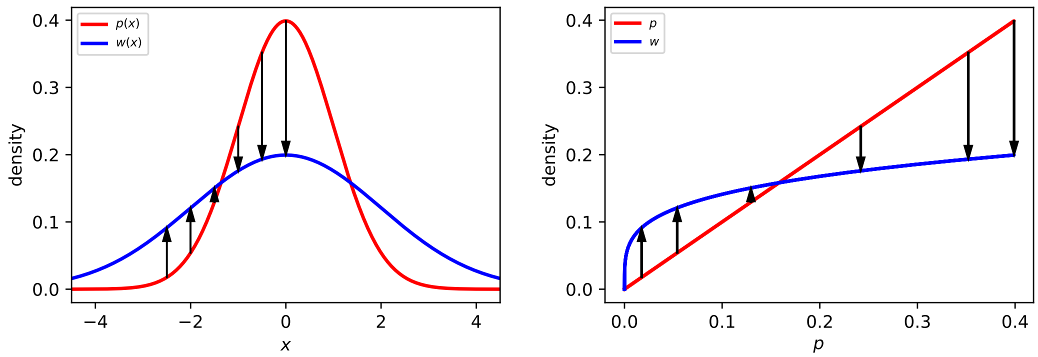

Figure 2: **Mapping density functions.** Left: probability density function (red) used by a DO; and weight density function (blue) used by a DM. The DO models X as a standard Gaussian random variable, $X \sim \mathcal{N}(0,1)$, whose density function is the red line. The DM models X as a Gaussian random variable with twice the scale, $X \sim \mathcal{N}(0,4)$, whose density function is the blue line. Comparing the two curves, the DM appears to the DO as someone who overestimates probability densities where they are low and underestimates them when high, indicated by vertical arrows. Right: the difference between weight and probability densities can be expressed by directly plotting, for any value of x, the weight density *vs.* the probability density observed at x. The arrows on the left correspond to the same x-values as on the right. Therefore, they start and end at identical vertical positions as on the left.

In the Gaussian case, we can write the density functions explicitly as

$$w(x) = \frac{1}{\sqrt{2\pi(\alpha\sigma)^2}} \exp\left[\frac{-(x-\mu)^2}{2(\alpha\sigma)^2}\right] \ , \tag{5}$$

and

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \ . \tag{6}$$

Eliminating $(x-\mu)^2$ from (Eq. 5) and (Eq. 6) yields an expression for the weight density as a function of the probability density,

$$w(p) = p^{\frac{1}{\alpha^2}} \frac{(2\pi\sigma^2)^{\frac{1-\alpha^2}{2\alpha^2}}}{\alpha} \ . \tag{7}$$

This is the blue curve in the right panel of Fig. 2. As a sanity check, consider the shape of $w(p)$: for a given value of α, it is just a power law in p with some prefactor that ensures normalisation. $\alpha > 1$ means that the DM uses a greater standard deviation than the DO. In this case, the exponent of p satisfies $\frac{1}{\alpha^2} < 1$, and the blue curve is above the diagonal for small densities and below it for large densities.

If instead $0 < \alpha < 1$, then the DM uses a smaller standard deviation than the DO. This might occur, for example, if the DM is an "insider" with better information than the DO. It can also occur if the DO observes two DMs and expresses the inferred densities as a function of the density with the higher relative uncertainty of the two. Instead of an inverse-S shape, both situations would lead to an S-shaped curve, see Appendix A.

Alternatively, we can express the identical information about the difference between models by plotting the CDFs, F_w and F_p. We do this in Fig. 3, where the inverse-S emerges purely from the DM's greater assumed scale, $\alpha\sigma$.

2.2 Different scales and locations

In Fig. 4 we explore what happens if both the scales and the locations of the DO's and DM's models differ. Visually, this produces an excellent fit to empirical data reported by Preston and Baratta (1948). A difference in assumed scales and locations of simple Gaussian distributions is sufficient to reproduce the observations. This suggests a different nomenclature and a conceptual clarification. The inverse-S curve does not mean that "probabilities are re-weighted." It means only that experimenters and their subjects have different views about appropriate models of, and responses to, a situation.[3]

[3] A shift in location only, as in Fig. 4 (top left), corresponds to a DM believing that outcomes are either consistently smaller (shift to the left) or larger (shift to the right) than implied by the DO's model. Price and Jones (2020) employ similar shifts of location to express optimism or pessimism of DMs in a rank-dependent expected utility model, where pessimism is protective against rare disastrous events.

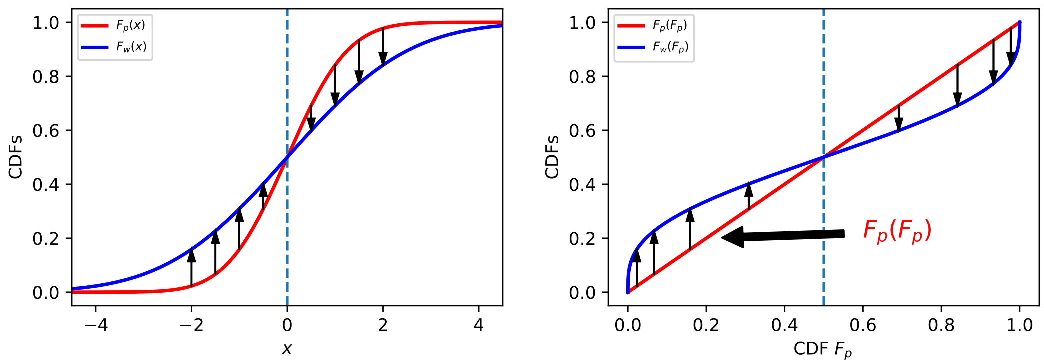

Figure 3: **Mapping CDFs.** Left: the DO models X as a standard Gaussian random variable, $X \sim \mathcal{N}(0, 1)$, with CDF, $F_p(x) = \Phi_{0,1}(x)$. The DM is more cautious and uses a wider Gaussian random variable, $X \sim \mathcal{N}(0, 4)$, depicted by $F_w(x)$ (blue). Following the vertical arrows (left to right), we see that: for low values of x (small cumulative probabilities) the cumulative weights are larger than the cumulative probabilities, $F_w(x) > F_p(x)$; the curves coincide at $x = 0$; and for large values of x (large cumulative probabilities) the cumulative weights are smaller than the cumulative probabilities, $F_w(x) < F_p(x)$. Right: the same CDFs as on the left but now plotted against the cumulative probability, F_p. Trivially, the cumulative probability, F_p, plotted against itself is the diagonal, while the cumulative weight, F_w, now displays the characteristic inverse-S shape of probability weighting. The arrows start and end at the same vertical values as on the left.

2.3 Different shapes

Numerically, our procedure can be applied to arbitrary distributions. For a set of outcomes, $\{x_i\}$:

1. construct the list of cumulative probabilities assumed by the DO, $\{F_p(x_i)\}$;
2. construct the list of cumulative weights used by the DM, $\{F_w(x_i)\}$; and
3. plot $\{F_w(x_i)\}$ *vs.* $\{F_p(x_i)\}$.

Of course, the DM could assume a weight distribution whose shape differs from that of the DO's probability distribution. The inverse-S arises whenever a DM assumes a greater scale for a unimodal distribution. To illustrate the generality of the procedure, we carry it out in Fig. 5 using Student's t-distribution

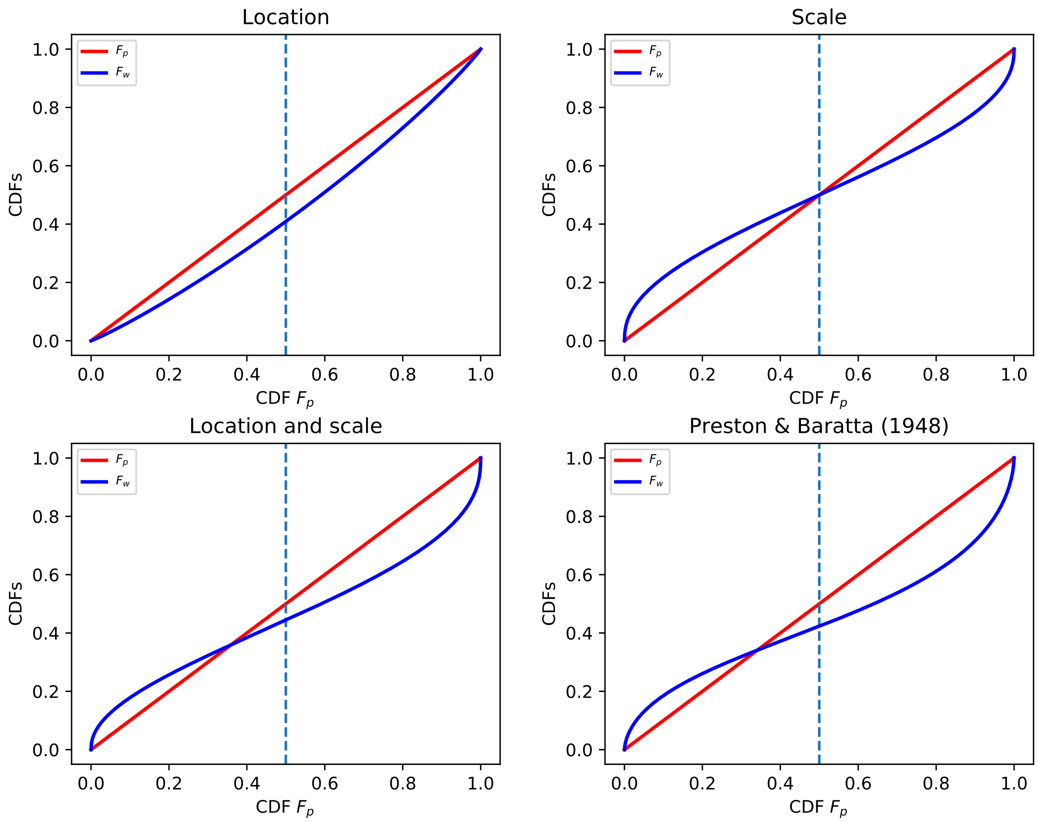

Figure 4: **CDF maps for Gaussian distributions.** Top left: difference in location. DO assumes location 0, scale 1; DM assumes location 0.23 (bigger than DO), scale 1. Top right: difference in scale. DO assumes location 0, scale 1; DM assumes location 0, scale 1.64 (broader than DO). Bottom left: differences in scale and location. DO assumes location 0, scale 1; DM assumes location 0.23 (bigger than DO), scale 1.64 (broader than DO). Bottom right: fit to observations reported by Preston and Baratta (1948). This is (Eq. 4) with $\gamma = 0.61$. Note the similarity to bottom left.

(hereafter t-distribution), with the DO and DM assuming different shape and location parameters.[4] The result is qualitatively similar to the bottom right panel of Fig. 3, corresponding to (Eq. 4).

[4]The density function of the t-distribution is

$$p\left(x\right) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\sigma}\left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}, \tag{8}$$

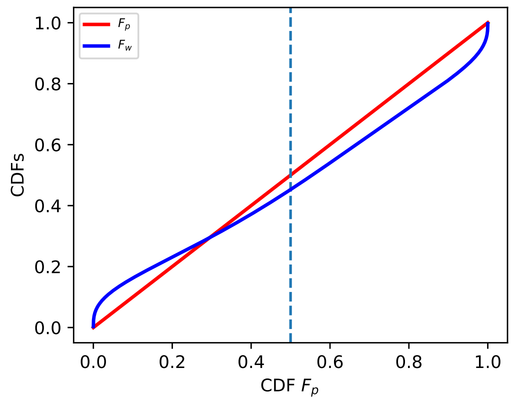

Figure 5: **Probability weighting for t-distributions.** The DM uses different shape and location parameters (1 and 0.35 respectively) from those of the DO (3 and 0.2).

3 Reasons for different models

Probability weighting is usually interpreted as a cognitive bias that leads to errors of judgement and poor decisions by DMs.[5]

We caution against this interpretation. At least we should keep in mind that it is unclear who suffers from the bias: experimenter or test subject (or neither or both). We are by no means the first to raise this question.

---

where ν is the shape parameter, σ is the scale parameter, and μ is the location parameter. The corresponding CDF is

$$F(x) = \begin{cases} 1 - \frac{1}{2} I_{\frac{\nu}{\left(\frac{x-\mu}{\sigma}\right)^2 + \nu}} \left(\frac{\nu}{2}, \frac{1}{2}\right) & \text{if } x - \mu \geq 0\,; \\ \frac{1}{2} I_{\frac{\nu}{\left(\frac{x-\mu}{\sigma}\right)^2 + \nu}} \left(\frac{\nu}{2}, \frac{1}{2}\right) & \text{if } x - \mu < 0\,, \end{cases} \tag{9}$$

where $I_x(a, b)$ is the incomplete beta function.

In the limit $\nu \to \infty$, the t-distribution converges to a Gaussian distribution with location μ and scale σ. We assume by default that $\sigma = 1$, so the t-distribution is effectively characterised by two parameters: shape, ν, and location, μ.

[5]Indeed, its originators presented it as such. Devising a gambling game played by human subjects, Preston and Baratta (1948, p. 183–184) argued that “the payment of amounts in excess of the mathematical expectation for the privilege of participation is peculiarly irrational” and claimed that “the failure of the price to correspond with the mathematical expectation is due to peculiarities of the player’s notion of the meaning of a given probability.”

Commenting on another so-called cognitive bias regarding probabilities, the representativeness fallacy, COHEN (1979) asked: "Whose is the fallacy?"

Whatever the answer, two observations are robust and interesting: first, disagreement is common; and, second, the disagreement tends to go in the same direction, with DMs assuming a greater range of plausible outcomes than DOs.

An explanation for the first observation is that probability is a slippery concept and the word is used to mean different things. This suggests that phrasing information about probabilities concretely should reduce disagreement between DO and DM. For example, the statement "10 out of 100 people have this disease" conveys more precise information than "the probability of having this disease is 0.1." Specifically, it tells us that a sample of people has been observed and what the size of the sample is.[6] Furthermore, GIGERENZER (2018) argues that statements involving integer counts, or what he calls natural frequencies, ("10 out of 100") are more readily understood by people than statements involving fractional probabilities ("0.1").[7]

The second observation may be explained as follows. A DO often has control over, and essentially perfect knowledge of, the decision problem he poses. A DM does not have such knowledge, and this ignorance will often translate into additional assumed uncertainty. For example, the DO may know the true probabilities of some gamble in an experiment, while the DM may have doubts about the DO's sincerity and her own understanding of the rules of the game. We will return to this in Sec. 3.2.

[6]This may contribute to an explanation of the observation of less overweighting in decisions-from-experience experiments reported in HERTWIG et al. (2004) and HERTWIG and EREV (2009).

[7]There is a growing body of literature on how to make cognitive biases disappear by changing the way probabilities are presented (GIGERENZER 1991) or by controlling for estimation errors and other confounders, for an example from social cognition see GALESIC et al. (2012).

3.1 Some meanings of "probability"

Many thousands of pages have been written about the meaning of probability. We will not attempt a summary of the philosophical debate and instead highlight a few relevant points. Their commonality is that although behavioural models often contain probabilities as parameters, probabilities are not directly observable.

Frequency-in-an-ensemble interpretation of probability

Consider the simple probabilistic statement: "the probability of rain here tomorrow is 70%." Tomorrow only happens once, so one might ask: in 70% of what will it rain? The technical answer to this question is often: rain happens in 70% of the members of an ensemble of computer simulations, run by a weather service, of what may happen tomorrow. So one interpretation of "probability" is "relative frequency in a hypothetical ensemble of simulated possible futures." It is thus a statement about a model.

Frequency-over-time interpretation of probability

In some situations, the statement "the probability of rain here tomorrow is 70%" refers to the relative frequency over time. Before the advent of computer models in weather forecasting, people used to compare today's measurements (of, say, wind and pressure) to those from the past – weeks, months, or even years earlier. Forecasts were made on the assumption that the weather tomorrow would resemble the weather that had followed similar conditions in the historical record.

Rather than a statement about outcomes of an *in silico* model, the statement may thus be a summary of real-world observations over a long time.

Degree-of-belief interpretation of probability

No matter how "probability" relates to a frequentist physical statement, whether with respect to an ensemble of simultaneously possible futures or to a sequence of actual past futures, it also corresponds to a mental state of believing something with a degree of conviction: "I'm 70% sure it will rain tomorrow."

Again, probabilities are not observable. For our purpose it suffices to say that there's no guarantee that a probabilistic statement will be interpreted by the receiver (the DM) as it was intended by whoever made the statement (the DO).

3.2 Consistent differences between DO and DM

Estimation errors for probabilities

Let's assume that both the DO and the DM mean by "probability" the relative frequency of an event in an infinitely long time series of observations. Of course, real time series have finite length, so probabilities defined this way are model parameters and cannot actually be observed. But, from a real time series, we can estimate the best values to put into a model, by counting how often we see an event.

As the probability of an event gets smaller, so does the number of times we count it in a finite time series. If we want to say something about the uncertainty in this number, we can measure it – or imagine measuring it – in several time series to see how much it varies. The variations in count from one time series to another get smaller for rarer events, but the *relative* variations get larger, and so does the relative uncertainty in our estimate of probabilities. Take an extreme simplified example: asymptotically an event occurs in 0.1% of observations, and we have a time series of 100 observations. Around 99.5% of such time series will contain 0 or 1 events. Naïvely, then, we would estimate the probability as either 0 or 1%. In other words, we would estimate the event as either impossible or occurring ten times more frequently than it really would in a long series. However, if the event occurs 50% of the time asymptotically, then around 99.5% of time series would contain between 35 and 65 events, leading to a much smaller relative error in probability estimates.

A DM who must estimate probabilities from observations is well advised to account for this behaviour of uncertainties in her decision making. Specifically, the DM should acknowledge that, due to her lack of information, *prima facie* rare events may be rather more common than her data suggest, while common events, being revealed more often, are more easily characterised. In such circumstances, caution may dictate that the DM assign to rare events higher probabilities than her estimates, commensurate with her uncertainty in them. This would look like probability weighting to a DO and, indeed, would constitute a mechanistic reason for it.[8]

Formalising these thoughts, we find that so long as relative uncertainties are larger for rare events than for common events – which, generically, they are – then an inverse-S curve emerges. See Appendix B for a detailed discussion. Here we make a simple scaling argument and then check it with a simulation. For an asymptotic (or "true") probability density $p(x)$, the number of events $n(x)$ we see in the small interval $[x, x+\delta x]$ in a time series of T observations is proportional to $p(x)$, to δx, and to T. So we have $n(x) \sim p(x)\delta x T$, where we mean by $\sim$ "scales like." We also know that such counts, for example in the simple Poissonian case, are random variables whose uncertainties scale like $\sqrt{n(x)}$.

If we knew the asymptotic probability density $p(x)$, we could make an estimate of the count as

$$n(x) \approx p(x)\delta x T \pm \sqrt{p(x)\delta x T}\ . \tag{10}$$

We would write $\hat{n}(x) \equiv p(x)\delta x T$ as the estimate of $n(x)$ and $\varepsilon\left[\hat{n}(x)\right] \equiv \sqrt{p(x)\delta x T}$ as its uncertainty. Of course, this situation seldom applies, because usually we do not know $p(x)$.

Conversely, and more realistically, if we observe a count $n(x)$, then we can use the scaling $p(x) \sim n(x)/T\delta x$ to make an estimate of the asymptotic

[8]Interestingly, Kahneman and Tversky (1979, p. 281) made the same point, noting that "overestimation that is commonly found in the assessment of the probability of rare events" has the same effect on human decisions as probability weighting. Since they assumed that subjects in experiments adopt unquestioningly the stated probabilities, they argued that probability weighting was necessary to explain their observations. We make no such assumption here.

probability density as

$$p(x) \approx \frac{n(x)}{T\delta x} \pm \frac{\sqrt{n(x)}}{T\delta x} \ . \tag{11}$$

We write $\hat{p}(x) \equiv n(x)/T\delta x$ as the estimate of $p(x)$, and

$$\varepsilon\left[\hat{p}(x)\right] \equiv \frac{\sqrt{n(x)}}{T\delta x} = \sqrt{\frac{\hat{p}(x)}{T\delta x}} \tag{12}$$

as its uncertainty, which we have expressed in terms of the estimate itself.

The standard error, $\sqrt{\hat{p}(x)/T\delta x}$, in an estimated probability density shrinks as the density decreases. However, the relative error in the estimate is $1/\sqrt{\hat{p}(x)T\delta x}$, which grows as the event becomes rarer. This is consistent with our claim, that low densities come with larger relative errors, and constitutes the key message of this section. Errors in density estimates behave differently for low densities than for high densities: absolute errors are smaller for lower densities, but relative errors are larger.[9]

Let's assume that the DM is aware of the uncertainties in her estimates and, furthermore, that she does not like surprises. To avoid surprises, she adds the standard error to her estimate of the probability density, $\hat{p}(x)$, in order to construct what will appear to the DO as the weight density, $w(x)$. In effect, she constructs a reasonable worst case for each of her estimates. After normalising, this conservative strategy yields generically,

$$w(x) = \frac{\hat{p}(x) + \varepsilon\left[\hat{p}(x)\right]}{\int_{-\infty}^{\infty}\left(\hat{p}(s) + \varepsilon\left[\hat{p}(s)\right]\right) ds} \ , \tag{13}$$

and specifically, for the type of uncertainty we consider,

$$w(x) = \frac{\hat{p}(x) + \sqrt{\frac{\hat{p}(x)}{T\delta x}}}{\int_{-\infty}^{\infty}\left(\hat{p}(s) + \sqrt{\frac{\hat{p}(s)}{T\delta x}}\right) ds} \ . \tag{14}$$

Note that the cautionary correction term in (Eq. 14) is parametrised by

[9] This argument is true for any distribution. Hertwig et al. (2004, pp. 537–8) make the same argument for the special case of a binomial distribution.

$T\delta x$, which scales like the number of observations in $[x, x+\delta x]$. As $T\delta x$ grows large, the correction vanishes and both $w(x)$ and $\hat{p}(x)$ become consistent with the asymptotic probability density, $p(x)$. With perfect information, a DM need not adjust decisions to account for uncertainty.

Does our analysis, culminating in (Eq. 13) and (Eq. 14), reproduce the stylised facts of probability weighting, in particular the inverse-S curve? We check in two ways. First, analytically, by applying the DM's cautionary correction in (Eq. 14) directly to some reference distributions. Second, by simulating the DM compiling counts of outcomes drawn from reference distributions, from which she estimates probability densities and their uncertainties, used to construct the weight density. The simulation is meant to explore how noisy the effect is when a DM experiences only a single time series. The Python code is available at `bit.ly/lml-pw-code-dm-count`, and a Jupyter notebook can be loaded to manipulate the code in an online environment at `bit.ly/lml-pw-dm-count-b`. In both cases, we treat the DO as using the reference distribution to make his predictions of the DM's behaviour.

Figure 6 shows the resulting probability and weight densities and CDF mappings generated by setting $\hat{p}(x)$ in (Eq. 14) to be the density functions of a Gaussian distribution and a fat-tailed t-distribution. Inverse-S curves are found for both distributions and the effect is more pronounced for the fat-tailed distribution.

Figure 7 shows the results of a computer simulation of a DM who observes a series of realisations of either Gaussian or t-distributed random variables, which she counts into bins. In the simulation, a probability density, $\hat{p}(x)$, is estimated for each bin as $n(x)/T\delta x$ and its uncertainty, $\varepsilon\left[\hat{p}(x)\right]$, is obtained numerically as the standard deviation in each $\hat{p}(x)$ over 1000 parallel simulations. The DM's weight density is then obtained according to (Eq. 13). Again, inverse-S curves are found for both distributions, corroborating our scaling arguments.

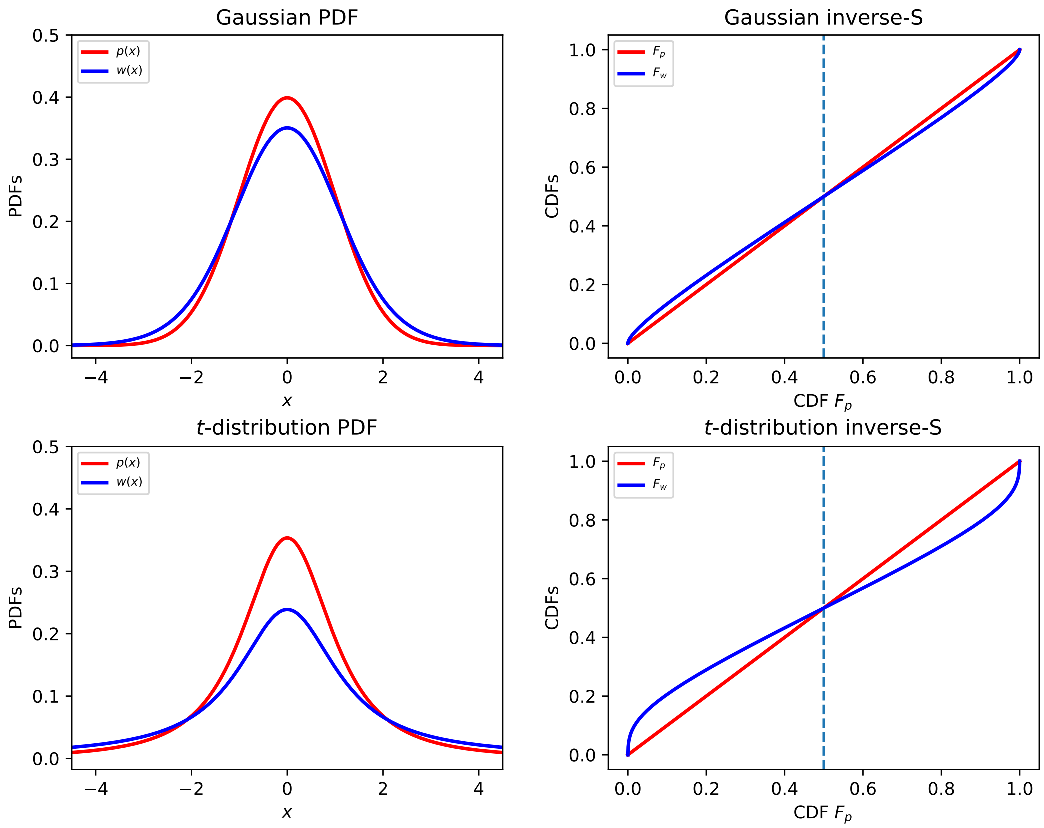

Figure 6: **Mapping density functions and CDFs with estimation errors.** Density functions (left) and inverse-S curves (right) arising when the DO assumes a Gaussian (scale 1, location 0, top line) or a t-distribution (shape 2, location 0, bottom line) and the DM constructs weight densities according to (Eq. 14) with $T\delta x = 10$. For the fat-tailed t-distribution (in the bottom line) the difference between $p(x)$ and $w(x)$ is more pronounced.

Typical situations of DO and DM: ergodicity

To recap, it is commonly observed that DOs assign lower weights to low probabilities than DMs. While a common interpretation is that the DM is wrong, we make no such judgement. In any decision problem, the aim of the decision must be taken into account. Crucially, this aim depends on the situation of the individual.

The two types of modellers (DO and DM) pursue different goals. In our

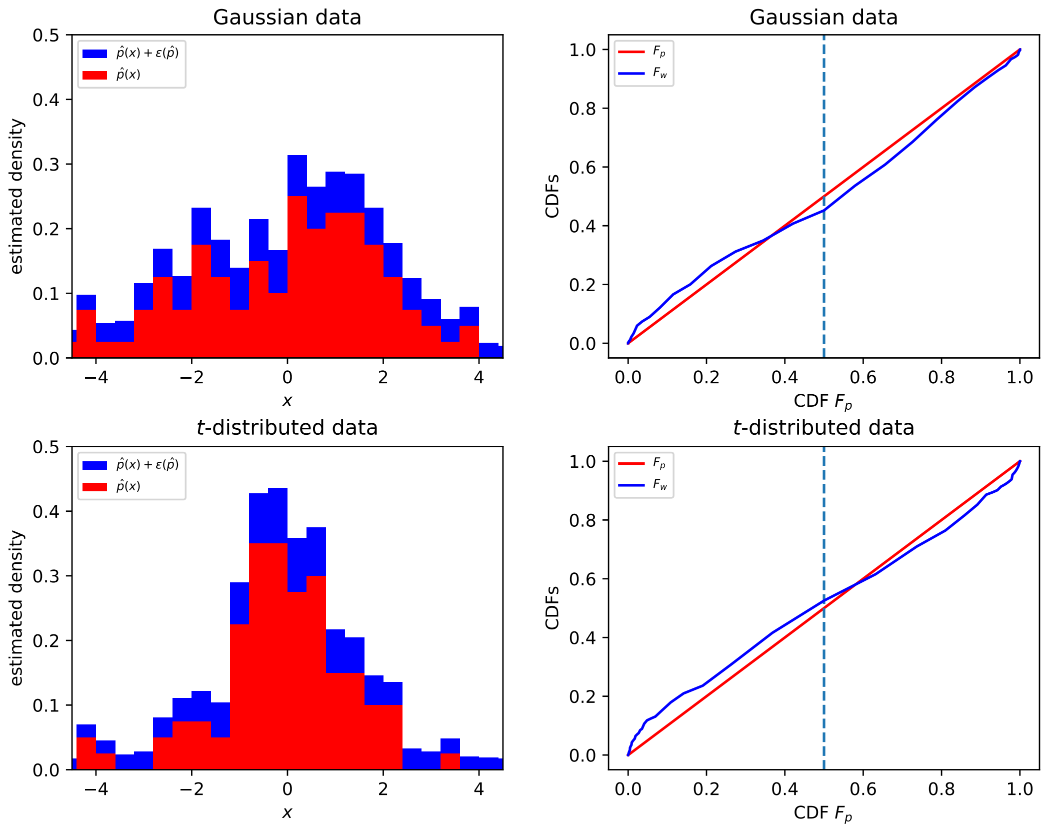

Figure 7: **Simulations of a DM estimating probability densities by counting events in a finite series of observations.** Left: estimated probability densities for $T = 100$ Gaussian (top; location 0, scale 2) and t-distributed (bottom; location 0, scale 1, shape 1.5) variates counted in bins of width $\delta x = 0.4$. Red bars show the estimates, $\hat{p}(x)$, and blue bars show the estimates with one standard error added, $\hat{p}(x) + \varepsilon\left[\hat{p}(x)\right]$. Right: inverse-S curves for a DO who assumes $p(x)$ follows a Gaussian (top) and t-distribution (bottom), while the DM uses weight density, $w(x)$, derived by normalising her conservative estimates (blue bars on left) according to (Eq. 13).

thought experiment, the DO is a behavioural scientist without personal exposure to the success or failure of the DM, whom we imagine as a test subject or someone whose behaviour is being observed in the wild. The DM, of course, has such exposure. Throughout the history of economics, it has been a common mistake, by DOs, to assume that DMs optimise what happens to them on average in an ensemble. To the DM, what happens to the ensemble is seldom a primary concern. Instead, she is concerned with what happens to her over time. Not distinguishing between these two perspectives is only permissible if they lead to identical predictions, meaning only if the relevant observables are ergodic (Peters 2019).

It is now well known that this is usually not the case in the following sense: DMs are usually observed making choices that affect their wealth, and wealth is usually modelled as a non-ergodic stochastic process. The ensemble average of wealth does not behave like the time average of wealth.

The most striking example is the universally important case of noisy multiplicative growth, the simplest model of which is geometric Brownian motion, $dx = x(\mu dt + \sigma dW)$. In the present context of human economic decisions, this is the most widely used model of the evolution of invested wealth. The average over the full statistical ensemble (often studied by the DO) of geometric Brownian motion grows as $\exp(\mu t)$. Each individual trajectory, on the other hand, grows in the long run as $\exp[(\mu - \frac{\sigma^2}{2})t]$. If the DO takes the ensemble perspective, he will deem the fluctuations irrelevant whereas, from the DM's time perspective, they reduce growth. So, while a DO curious about the ensemble may suffer no consequences from disregarding rare events, hedging against such events is central to the DM's success.

The difference between how these two perspectives evaluate the effects of probabilistic events is qualitatively in line with the observed phenomena we set out to explain. The DM typically has large uncertainties, especially for extreme events, and has an evolutionary incentive to err on the side of caution, *i.e.* to behave as though extreme events have a higher probability than in the DO's model.

4 Fitting the model to experimental results

Considering the noise in the data in Fig. 1, we might speculate that the inverse-S function of Tversky and Kahneman (1992), $F_w^{TK}(F_p)$ in (Eq. 4), would fit the data no better than the functions arising from our mechanistic model.

This is particularly evident in the bottom panels of Fig. 4, which show for Gaussian distributions that a weight density, $w(x)$, whose scale and location differ from those of the probability density, $p(x)$, reproduces the functional shape of $F_w^{TK}(F_p)$ fitted to the data in Preston and Baratta (1948).

For completeness and scientific hygiene, in the present section we fit location and scale parameters in the Gaussian and t models for F_w to experimental data from Preston and Baratta (1948) (depicted in Fig. 1) and from Tversky and Fox (1995). Specifically, in the Gaussian model we fit the location and scale parameters μ and σ in the CDF,

$$F_w(x) = \Phi\left(\frac{\Phi^{-1}(F_p(x)) - \mu}{\sigma}\right) , \tag{15}$$

where Φ is the CDF of the standard normal distribution. In the t-model, we fit the location and shape parameters, μ and ν, in the CDF, $F_w(x)$, of a t-distributed random variable (see Sec. 2.3). In both cases, we assume that $F_p(x)$ is that of a standard normal distribution.

In addition to $F_w^{TK}(F_p)$ in (Eq. 4), we fit the function

$$F_w^L(F_p; \delta, \gamma) = \frac{\delta F_p^\gamma}{\delta F_p^\gamma + (1 - F_p)^\gamma} , \tag{16}$$

suggested by Lattimore et al. (1992) to describe probability weighting parametrically (widely used, *i.a.* in Tversky and Wakker (1995) and Prelec (1998)). The reason for fitting (Eq. 16) is to ensure a fair comparison: the Gaussian and t-distribution models are characterised by two parameters, whereas (Eq. 4) only has one free parameter. Equation (16) has two parameters.

Figure 8 presents the results. We obtain good fits for both Gaussian and t-distributions, as well as for (Eq. 4) and (Eq. 16), to both sets of experimental data. It is practically impossible to distinguish between the fitted functions within standard errors. We conclude that our model fits the data well, and unlike (Eq. 4) or (Eq. 16), the fitted functions are directly derived from a physically plausible mechanism. They are not purely phenomenological.

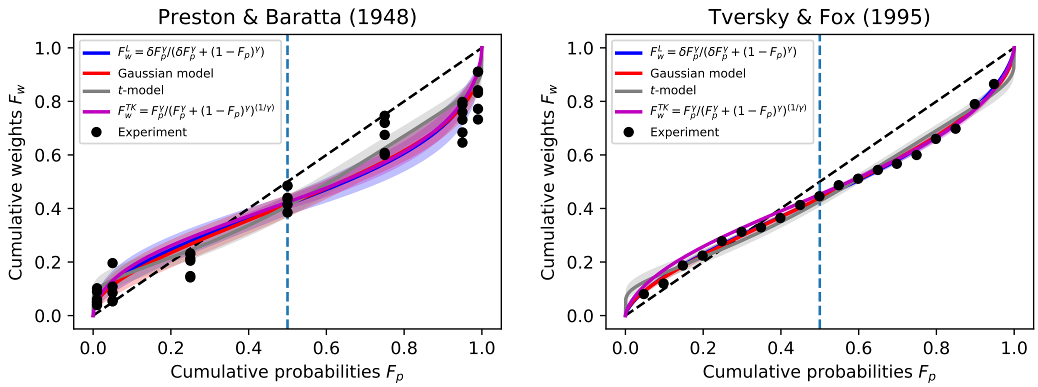

Figure 8: **Model fitting to experimental data from Preston and Baratta (1948) (left) and Tversky and Fox (1995) (right).** Left: Lattimore et al. (1992) (Eq. 16): $\delta = 0.71$ ($SE = 0.05$), $\gamma = 0.56$ (± 0.04); Gaussian model: $\mu = 0.36$ (± 0.07), $\sigma = 1.65$ (± 0.10); t-model: $\nu = 1.24$ (± 0.20), $\mu = 0.28$ (± 0.06); Tversky and Kahneman (1992) (Eq. 4): $\gamma = 0.61$ (± 0.02). Right: Lattimore et al. (1992): $\delta = 0.77$ (± 0.01), $\gamma = 0.69$ (± 0.01); Gaussian model: $\mu = 0.22$ (± 0.01), $\sigma = 1.41$ (± 0.03); t-model: $\nu = 1.41$ (± 0.21), $\mu = 0.22$ (± 0.03); Tversky and Kahneman (1992): $\gamma = 0.68$ (± 0.01). Shaded areas indicate two standard errors in the fitted parameter values. The fit was done by implementing the Levenberg-Marquardt algorithm (Levenberg 1944) for nonlinear least squares curve fitting.

5 Discussion

On 28 February 2020, an opinion piece appeared on Bloomberg (Sunstein 2020), arguing that people's fears about a potential coronavirus outbreak in the United States were attributable to an extreme case of probability weighting: people were neglecting the fact that such an event had a low probability. When it was published, many disagreed: it seemed reasonable to them to take precautions; the author himself, they said, might have underestimated both the severity and likelihood of a pandemic. In effect, they asked the Cohen (1979) question: "Whose is the fallacy?"

As we write this, in December 2020, more than 300,000 Americans have died of the virus.

This episode illustrates that an inverse-S curve is indeed a neutral indicator of a difference in opinion. It says nothing about who is right and who is wrong and it calls for caution against the cavalier presumption of behavioural bias. Assuming other people's judgments are systematically flawed can lead us to disregard the possibility that we ourselves may be wrong. Taking our own uncertainty into account is crucial when the precautionary principle applies: if we are unsure whether a snake in the grass is poisonous, we had better treat it as poisonous and stay away.

The term "probability weighting" suggests an obscure mental process, in which a DM carries out operations on probabilities. It seems more natural to us to consider a DM modelling events about whose probabilities she is unsure. From this latter point of view, it is easy to think of reasons for a DM's model to differ from a DO's. DMs will often have cause to include additional uncertainty, leading to the frequently observed inverse-S curve.

The model of estimating probabilities from real time series, which we discuss in Sec. 3, has qualitative features that display a degree of universality. Relative errors in a DM's probability estimates are always greater for rarer events. A dislike of the unexpected, which explains the systematic overestimation of low probabilities, is similarly common. Probability weighting, as classically presented, is purely descriptive and comes with the unwarranted connotation of DMs suffering from cognitive errors. The same phenomenon may be well characterised as DMs making wise decisions given the information available to them. Such information is necessarily limited because, for example, DMs are constrained to collect such information over time.

References

Adamou, Alexander, Yonatan Berman, Diomides Mavroyiannis, and Ole Peters (2020). “Microfoundations of Discounting”. *arXiv:1910.02137*. arXiv: 1910.02137 (cit. on p. 7).

Berman, Yonatan and Mark Kirstein (2020). “Risk Preferences in Time Lotteries”. *researchers.one*. url: bit.ly/TimeLotteries (cit. on p. 7).

Bernoulli, Daniel (1738). “Specimen Theoriae Novae de Mensura Sortis. Translation “Exposition of a New Theory on the Measurement of Risk” by L. Sommer (1954)”. *Econometrica* 22.1, 23–36. doi: 10.2307/1909829. url: http://www.jstor.org/stable/1909829 (cit. on p. 5).

Cohen, L. Jonathan (1979). “On the psychology of prediction: Whose is the fallacy?” *Cognition* 7 (4), 385–407. doi: 10.1016/0010-0277(79)90023-4 (cit. on pp. 14, 24).

Erev , Ido, Eyal Ert, Alvin E. Roth, Ernan Haruvy, Stefan M. Herzog, Robin Hau, Ralph Hertwig, Terrence Stewart, Robert West, and Christian Lebiere (2010). “A choice prediction competition: Choices from experience and from description”. *Journal of Behavioral Decision Making* 23 (1), 15–47. doi: 10.1002/bdm.683 (cit. on p. 5).

Fox, Craig R. and Liat Hadar (2006). ““Decisions from experience” = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004)”. *Judgement and Decision Making* 1 (2). url: http://journal.sjdm.org/06144/jdm06144.htm (cit. on p. 6).

Galesic, Mirta, Henrik Olsson, and Jörg Rieskamp (2012). “Social Sampling Explains Apparent Biases in Judgments of Social Environments”. *Psychological Science* 23 (12), 1515–1523. doi: 10.1177/0956797612445313 (cit. on p. 14).

Gigerenzer, Gerd (1991). “How to Make Cognitive Illusions Disappear: Beyond “Heuristics and Biases””. *European Review of Social Psychology* 2 (1), 83–115. doi: 10.1080/14792779143000033 (cit. on p. 14).

Gigerenzer, Gerd (2018). “The Bias Bias in Behavioral Economics”. *Review of Behavioral Economics* 5 (3-4), 303–336. doi: 10.1561/105.00000092 (cit. on p. 14).

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. doi: 10.1111/j.0956-7976.2004.00715.x (cit. on pp. 5, 6, 14, 18).

Hertwig, Ralph and Ido Erev (2009). “The description–experience gap in risky choice”. *Trends in Cognitive Sciences* 13 (12), 517–523. doi: 10.1016/j.tics.2009.09.004 (cit. on pp. 5, 14).

Kahneman, Daniel and Amos Tversky (1979). “Prospect Theory: An Analysis of Decision under Risk”. *Econometrica* 47 (2), 263–291. doi: 10.2307/1914185 (cit. on pp. 2, 5, 17).

Laplace, Pierre Simon (1814). *Théorie analytique des probabilités*. 2nd ed. Paris, Ve. Courcier (cit. on p. 5).

Lattimore, Pamela K., Joanna R. Baker, and A. Dryden Witte (1992). “Influence of Probability on Risky Choice: A Parametric Examination”. *Journal of Economic Behavior and Organization* 17 (3), 377–400. doi: 10.1016/S0167-2681(95)90015-2 (cit. on pp. 23, 24).

Levenberg, Kenneth (1944). “A Method for the Solution of Certain Non-Linear Problems in Least Squares”. *Quarterly of Applied Mathematics* 2 (2), 164–168. doi: 10.1090/qam/10666 (cit. on p. 24).

Martins, André C. R. (2006). “Probability biases as Bayesian inference”. *Judgement and Decision Making* 1 (2), 108–117. url: http://journal.sjdm.org/vol1.2.htm (cit. on p. 6).

MEDER, DAVID, FINN RABE, TOBIAS MORVILLE, KRISTOFFER H. MADSEN, MAGNUS T. KOUDAHL, RAY J. DOLAN, HARTWIG R. SIEBNER, and OLIVER J. HULME (2019). "Ergodicity-Breaking Reveals Time Optimal Economic Behavior in Humans". *arXiv:1906.04652*. arXiv: 1906.04652 (cit. on p. 7).

VON NEUMANN, JOHN and OSKAR MORGENSTERN (1944). *Theory of Games and Economic Behavior*. Princeton University Press (cit. on p. 5).

PETERS, OLE (Nov. 2011a). "Optimal leverage from non-ergodicity". *Quantitative Finance* 11.11, 1593–1602. DOI: 10.1080/14697688.2010.513338. URL: http://dx.doi.org/10.1080/14697688.2010.513338 (cit. on p. 7).

PETERS, OLE (Dec. 2011b). "The time resolution of the St Petersburg paradox". *Phil. Trans. R. Soc. A* 369.1956, 4913–4931. DOI: 10.1098/rsta.2011.0065. URL: http://dx.doi.org/10.1098/rsta.2011.0065 (cit. on p. 7).

PETERS, OLE (2019). "The Ergodicity Problem in Economics". *Nature Physics* 15 (12), 1216–1221. DOI: 10.1038/s41567-019-0732-0 (cit. on pp. 6, 7, 22).

PRELEC, DRAŽEN (1998). "The Probability Weighting Function". *Econometrica* 66 (3), 497–527. DOI: 10.2307/2998573 (cit. on p. 23).

PRESTON, MALCOLM G. and PHILIP BARATTA (1948). "An Experimental Study of the Auction-Value of an Uncertain Outcome". *American Journal of Psychology* 61 (2), 183. DOI: 10.2307/1416964 (cit. on pp. 2–5, 7, 10, 12, 13, 23, 24).

PRICE, MICHAEL HOLTON and JAMES HOLLAND JONES (2020). "Fitness-Maximizers Employ Pessimistic Probability Weighting for Decisions Under Risk". *Evolutionary Human Sciences* 2, e28. DOI: 10.1017/ehs.2020.28 (cit. on p. 10).

Savage, Leonard J. (1954). *The Foundations of Statistics*. Dover Publications, New York (cit. on p. 5).

Seo, Daewon, Ravi Kiran Raman, Joong Bum Rhim, Vivek K Goyal, and Lav R. Varshney (2019). "Beliefs in Decision-Making Cascades". *IEEE Transactions on Signal Processing* 67 (19), 5103–5117. doi: 10.1109/tsp.2019.2935865 (cit. on p. 6).

Stewart, Neil, Stian Reimers, and Adam J. L. Harris (2015). "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. doi: 10.1287/mnsc.2013.1853 (cit. on p. 5).

Sunstein, Cass R. (2020). "The Cognitive Bias That Makes Us Panic About Coronavirus". url: https://www.bloomberg.com/opinion/articles/2020-02-28/coronavirus-panic-caused-by-probability-neglect (visited on 04/04/2020) (cit. on p. 24).

Tversky , Amos and Craig R. Fox (1995). "Weighing Risk and Uncertainty". *Psychological Review* 102 (2), 269–283. doi: 10.1037/0033-295X.102.2.269 (cit. on pp. 23, 24).

Tversky , Amos and Daniel Kahneman (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty". *Journal of Risk and Uncertainty* 5 (4), 297–323. doi: 10.1007/BF00122574 (cit. on pp. 2, 3, 5, 7, 22, 24).

Tversky , Amos and Peter Wakker (1995). "Risk Attitudes and Decision Weights". *Econometrica* 63 (6), 1255–1280. doi: 10.2307/2171769 (cit. on p. 23).

Ungemach, Christoph, Nick Chater, and Neil Stewart (2009). "Are Probabilities Overweighted or Underweighted When Rare Outcomes Are Experienced (Rarely)?" *Psychological Science* 20 (4), 473–479. doi: 10.1111/j.1467-9280.2009.02319.x (cit. on p. 6).

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), 140–176. doi: 10.1037/bul0000115 (cit. on p. 6).

A Cases of a regular S

We saw how the robust qualitative observation of the inverse-S curve in Fig. 1 arises when the DM uses a larger scale in her model of uncertainty than the DO. In the Gaussian case, it was possible to express the weight density, w, as a function of the probability density, p,

$$w(p) = p^{\frac{1}{\alpha^2}} \frac{\left(2\pi\sigma^2\right)^{\frac{1-\alpha^2}{2\alpha^2}}}{\alpha} , \tag{17}$$

where α is the ratio of the DM's scale parameter to the DO's.

Suppose instead that the DM uses a smaller scale than the DO, such that $0 < \alpha < 1$. In other words, the DM is more certain about the value X will take than the DO. We expect this to happen if, for instance, the DM has insider information. Then the CDF mapping $F_w(F_p)$ assumes the shape of an S, rather than an inverse-S, as shown in Fig. 9 (compare with Fig. 2 and Fig. 3 for $\alpha > 1$). The prediction of an S curve under such conditions is an empirically testable prediction with discriminating power: it does not arise under the classical interpretation of probability weighting.

Another common case might be when there is no "true" model, *e.g.*, if the DO observes the choices of two DMs and expresses the implied densities as a function of the DM's whose model has relatively more uncertainty.

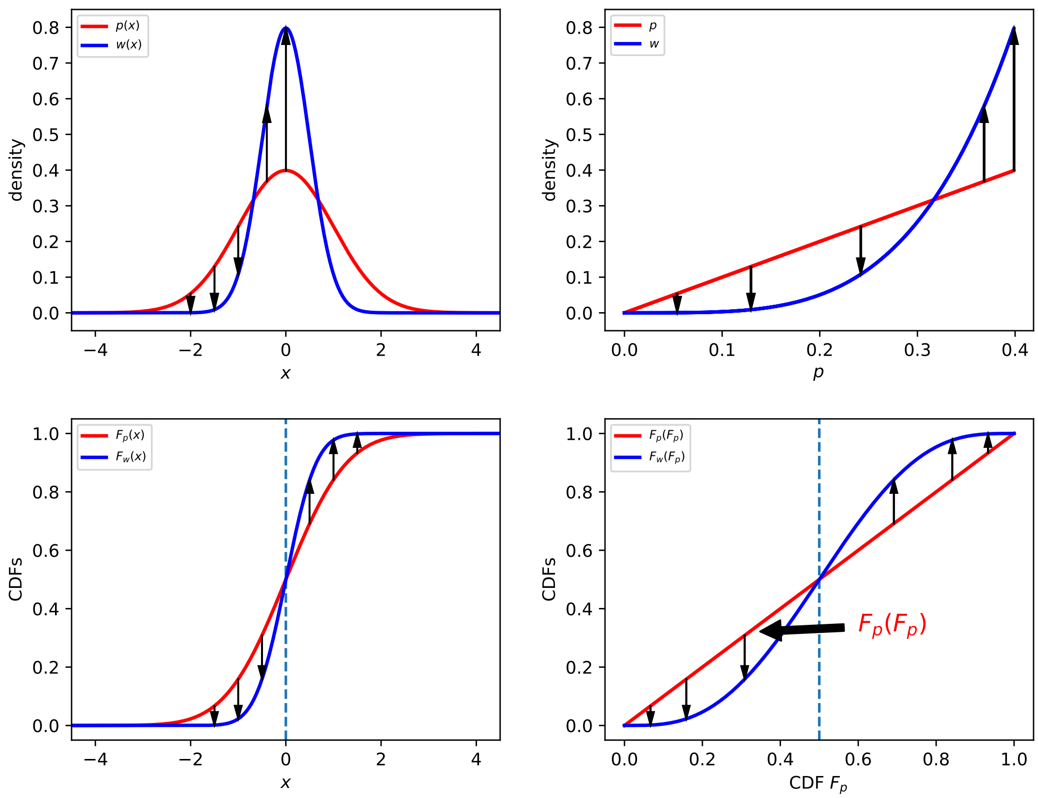

Figure 9: **Generating an S curve.** Analogues of Fig. 2 and Fig. 3 for the case $0 < \alpha < 1$.

B Inverse-S from relative errors in probabilities

For an inverse-S curve to emerge, small probability densities have to be overestimated ($w > p$) and large ones underestimated ($w < p$), as is indeed the case, for example in Fig. 6. Let us connect this statement to one about relative uncertainties. The weight density is constructed by adding the probability density $p(x)$ to its uncertainty $\varepsilon\,[p(x)]$ and normalising, as we did in (Eq. 13), *i.e.*

$$w(x) = \frac{p(x) + \varepsilon\,[p(x)]}{\int_{-\infty}^{\infty} (p(s) + \varepsilon\,[p(s)])\, ds} \, . \tag{18}$$

This can be expressed as

$$w(x) = p(x) \left(\frac{1 + \frac{\varepsilon[p(x)]}{p(x)}}{\int_{-\infty}^{\infty} p(s) \left\{ 1 + \frac{\varepsilon[p(s)]}{p(s)} \right\} ds} \right) \quad , \tag{19}$$

where $\frac{\varepsilon[p(x)]}{p(x)}$ is the relative error, and the denominator of (Eq. 19) is a normalisation constant. If the relative error is larger for smaller densities, then smaller densities are enhanced more (the summand $\frac{\varepsilon[p(x)]}{p(x)}$ in the numerator is greater) than larger densities. The normalisation constant scales down all densities equally and, where the enhancement was greater, $w(x)$ ends up above $p(x)$. Where it was lower, $w(x)$ ends up below $p(x)$. So, if the relative error is larger for smaller densities, an inverse-S curve emerges.

We can say one more thing about this procedure. If an inverse-S curve exists, then $p(x)$ and $w(x)$ cross somewhere, see Fig. 6. This happens when the relative error attains its expectation value (with respect to the density p). Rewriting (Eq. 19) as

$$w(x) = p(x) \left(\frac{1 + \frac{\varepsilon[p]}{p}}{1 + \left\langle \frac{\varepsilon[p]}{p} \right\rangle} \right) , \tag{20}$$

we see that $w(x) = p(x)$ when $\frac{\varepsilon[p]}{p} = \left\langle \frac{\varepsilon[p]}{p} \right\rangle$.