What are we weighting for?

A mechanistic model for probability weighting*

Ole Peters[†1,2], Alexander Adamou[‡1], Mark Kirstein[§1,3,4], and Yonatan Berman[¶1,5]

[1] *London Mathematical Laboratory, UK*
[2] *Santa Fe Institute, NM, USA*
[3] *Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany*
[4] *Institute of Mathematics, Leipzig University, Germany*
[5] *Stone Center on Socio-Economic Inequality, City University of New York, NY, USA*

October 2, 2020

Abstract

Behavioural economics provides labels for patterns in human economic behaviour. Probability weighting is one such label. It expresses a mismatch between probabilities used in a formal model of a decision (*i.e.* model parameters) and probabilities inferred from real people's decisions (the same parameters estimated empirically). The inferred probabilities are called "decision weights." It is considered a robust experimental finding that decision weights are higher than probabilities for rare events, and (necessarily, through normalisation) lower than probabilities for common events. Typically this is presented as a cognitive bias, *i.e.* an error of judgement by the person. Here we point out that the same observation can be described differently: broadly speaking, probability weighting means that a decision maker has greater uncertainty in his model than the observer. We offer a plausible mechanism whereby such differences in uncertainty arise naturally: when a decision maker must estimate probabilities as frequencies in a time series while the observer knows them *a priori*. This suggests an alternative presentation of probability weighting as a principled response by a decision maker to uncertainties unaccounted for in an observer's model.

Keywords Decision Theory, Prospect Theory, Probability Weighting, Ergodicity Economics

JEL Codes C61 · D01 · D81

*The latest version of the manuscript is available at bit.ly/lml-pw. The code to generate the figures and simulations is available at bit.ly/lml-pw-code.

† ✉ o.peters@lml.org.uk
‡ ✉ a.adamou@lml.org.uk
§ ✉ m.kirstein@lml.org.uk
¶ ✉ y.berman@lml.org.uk

1 Introduction

Probability weighting is a concept that originated in prospect theory (Kahneman and Tversky 1979; Tversky and Kahneman 1992). It is one way to conceptualise a pattern in human behaviour of caution with respect to formal models. This is best explained by a thought experiment, in which

- a *disinterested observer* (DO), such as an experimenter, tells
- a *decision maker* (DM)

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 optimization) in which the DM uses a probability that differs systematically from what the DO has declared. The apparent probabilities, inferred from the DM's decisions, are called "*decision weights*." We will adopt this nomenclature here.

- By "*probabilities,*" expressed as probability density functions (PDFs) and denoted $p(x)$, we will mean the numbers specified by a DO.
- By "*decision weights,*" also expressed as PDFs and denoted $w(x)$, we will mean the numbers that best describe the behaviour of a DM in the DO's behavioural model.[1]

Here, x is a realisation of a random variable, X. For example, X might be the payout of a gamble which the DM is invited to accept or decline.

Different behavioural models may result in different inferred decision weights. Our focus is not on how these weights are inferred, but on the robust observation that decision weights, $w(x)$ (used by DMs), are higher than probabilities, $p(x)$ (declared by DOs), for extreme events, *i.e.* when $p(x)$ is small. Thus, we do not consider any specific behavioural model: our aim is to find a general mechanistic explanation to probability weighting.

Probability weighting is often summarised visually by comparing

- cumulative density functions (CDFs) for probabilities, denoted

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

- and CDFs for decision weights, denoted

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

In Fig. 1 we reproduce the first such visual summary from Tversky and Kahneman (1992, p. 310).

Plotting F_w as a function of F_p generates a curve, whose generic shape we shall call the "inverse-S". The inverse-S is the main observational finding in probability weighting: it sits above the diagonal $F_w = F_p$ for events of low cumulative probability (such that

[1] In the literature, decision weights are not always normalised, but for simplicity we will work with normalised decision weights. Mathematically speaking, they are therefore proper probabilities even though we don't call them that. Our results are unaffected because normalising just means dividing by a constant (the sum or integral of the non-normalised decision weights).

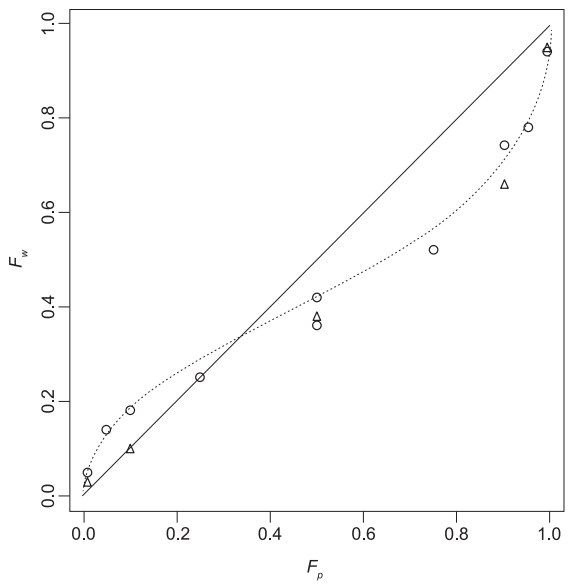

Figure 1: **Empirical phenomenon of probability weighting.** Cumulative decision weights F_w (used by decision makers) versus cumulative probabilities F_p (used by disinterested observers), as reported by Tversky and Kahneman (1992, p. 310, Fig. 1, relabelled axes). The figure is to be read as follows: pick a point along the horizontal axis (the cumulative probability F_p used by a DO) and look up the corresponding value on the vertical axis of the dotted inverse-S curve (the cumulative decision weight F_w used by a DM). Low cumulative probabilities (left) are exceeded by their corresponding cumulative decision weights, and for high cumulative probabilities it's the other way around. It's the inverse-S shape of the curve that indicates this qualitative relationship.

$F_w > F_p$ 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 will use the terms *location*, *scale*, and *shape* when discussing probability distributions. Consider a standard normal distribution $\mathcal{N}(0, 1)$, whose parameters indicate location 0 and squared scale 1. (For a Gaussian, the location is the mean and scale is the standard deviation.) 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 identically-shaped distribution, but with location 0 and scale 1:

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

Thus the PDF of Z, $p(z)$, is a density function with location $\mu_Z = 0$ and scale $\sigma_Z = 1$. In a graph of a distribution, a change of location shifts the curve to the left or right, and a change in scale shrinks or blows up the width of its features. Neither operation changes the *shape* of the distribution: two distributions 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 was founded 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 also failed to fit some observations, so prospect theory (Kahneman and Tversky 1979) and then cumulative prospect theory (CPT) (Tversky and Kahneman 1992) were proposed as descriptive decision theories. CPT adds a distortion of probabilities to EUT's distortion of monetary amounts. This distortion of probabilities is known as probability weighting and is the focus of this paper. 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.[2]

As research continued, critiques of CPT and specifically of probability weighting emerged. With both monetary amounts and probabilities now transformed by freely chosen nonlinear functions, CPT is extremely flexible and prone to 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 decision 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.[3] 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 questions of statistical significance. 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, see 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 decision 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 af-

[2] An early example of this interpretation is in Preston and Baratta (1948, Fig. 1, p. 188).

[3] 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 decisions-from-experience paradigm is feasible.

ter 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.

In our own work, ergodicity economics, the DM's experience of outcomes along a single timeline is fundamental (Peters 2019). Therefore, in this paper we ask 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, ergodicity economics rejects EVT because there is no physical reason for expectation values to reflect what happens to an individual over time. Instead ergodicity economics uses as its null model the optimization of appropriate growth rates evaluated in the long-time limit (Peters 2011a,b, 2019). For clarity, this is in contrast to conventional decision theories, including CPT, which maintain expectation values (suitably distorted) as their optimands. Ergodicity economics 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 differences between models of uncertainty exist. The mechanism we propose can generate mismatches between probabilities and decision weights which produce a variety of probability weighting curves, including the standard inverse-S curve of CPT. The inverse-S is predicted whenever a DM cautiously incorporates more uncertainty into his model than a DO, which we imagine to be a commonly encountered scenario in experiments. More specifically, we find the inverse-S when a DM cautiously adjusts a frequentist probability estimate to account for sampling errors.

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 model of the observable. Nor do we assume the DM to know "the truth" – he has another model of the observable. 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

2.1.1 Tversky and Kahneman

Tversky and Kahneman (1992) chose to fit the empirical data in Fig. 1 with the following function,

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

which maps from one CDF, F_p, to another, F_w. We note that no mechanistic motivation was given for fitting this specific family of CDF mappings, parametrised by γ. The motivation is purely phenomenological: with $\gamma < 1$, this function can be made to fit to the data reasonably well.[4]

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, F_w^{TK} has the following property: any curvature moves the intersection with the diagonal away from the mid-point 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. We consider the key observation to be the inverse-S shape, whereas the shift to the left may be an artefact of the function chosen for the fit.

2.1.2 Scale, location, and the inverse-S

We now make explicit how the robust qualitative observation of the inverse-S shape in Fig. 1 emerges when the DM uses a larger scale in his model of the observable than the DO.

We illustrate this with a Gaussian distribution. Let's assume that a DO models an observable x – which will often be a future change in wealth – as a Gaussian random variable with location μ and variance σ^2. And let's further assume that a DM models the same observable as a Gaussian random variable with the same location, μ, but with a greater scale, so that the variance is $(\alpha\sigma)^2$. 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 his model, then decision weights are higher than probabilities for low-probability events, and (because of normalisation) lower than probabilities for high-probability events. We can express this by plotting, for any value of x, the decision weight *vs.* the probability observed at x, right panel of Fig. 2.

In the Gaussian case we can write the distributions 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 the following expression for decision weight as a function of probability:

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

We plot this in the right panel of Fig. 2. As a sanity check, consider the shape of the $w(p)$ (blue curve, right panel Fig. 2): for a given value of α, it is just a power law in p with some pre-factor that ensures normalization. $\alpha > 1$ 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. $0 < \alpha < 1$

[4]The mappings like F^{TK} in (Eq. 4) or later F^L in (Eq. 16) from one CDF to another CDF are themselves not CDFs. We denote the difference between such mappings and proper CDFs which depend on x, like F_p and F_w in (Eq. 1) and (Eq. 2) resp., by a superscript throughout the paper.

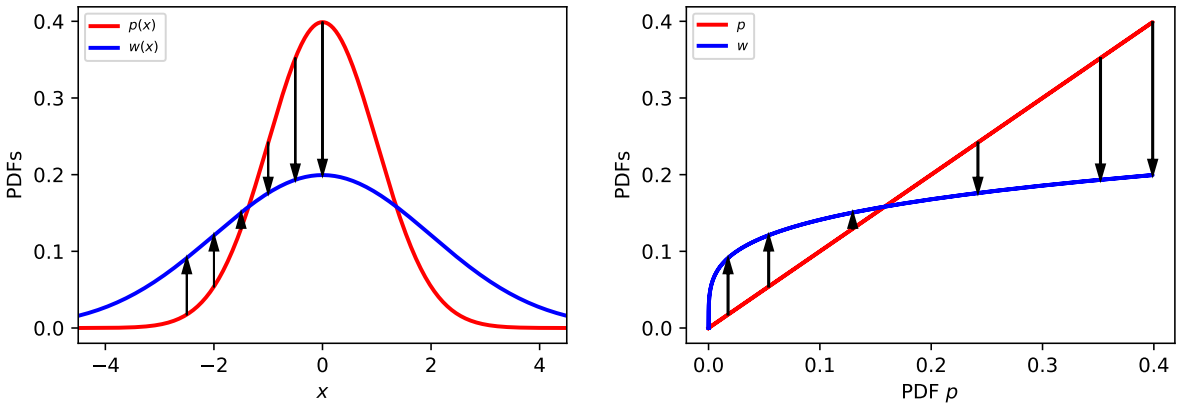

Figure 2: **Mapping PDFs.** Left: probability PDF (red), estimated by a DO; and decision-weight PDF (blue), estimated by a DM. The DO models x with a best estimate for the scale (standard deviation) and assumes the true frequency distribution is the red line. The DM models x with a greater scale (here 2 times greater, $\alpha = 2$), and assumes the true frequency distribution is the blue line. Comparing the two curves, the DM appears to the DO as someone who over-estimates probabilities of low-probability events and underestimates probabilities of high-probability events, indicated by vertical arrows. Right: the difference between decision weights and probabilities can also be expressed by directly plotting, for any value of x, the decision weight *vs.* the probability observed at x. This corresponds to a non-linear distortion of the horizontal axis. The arrows on the left correspond to the same x-values as on the right. They therefore start and end at identical vertical positions as on the left. Because of the non-linear distortion of the horizontal axis, they are shifted to different locations horizontally.

means that 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. Instead of an inverse-S shape, this would lead to an S-shaped curve, see Appendix B.

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, $\alpha\sigma$.

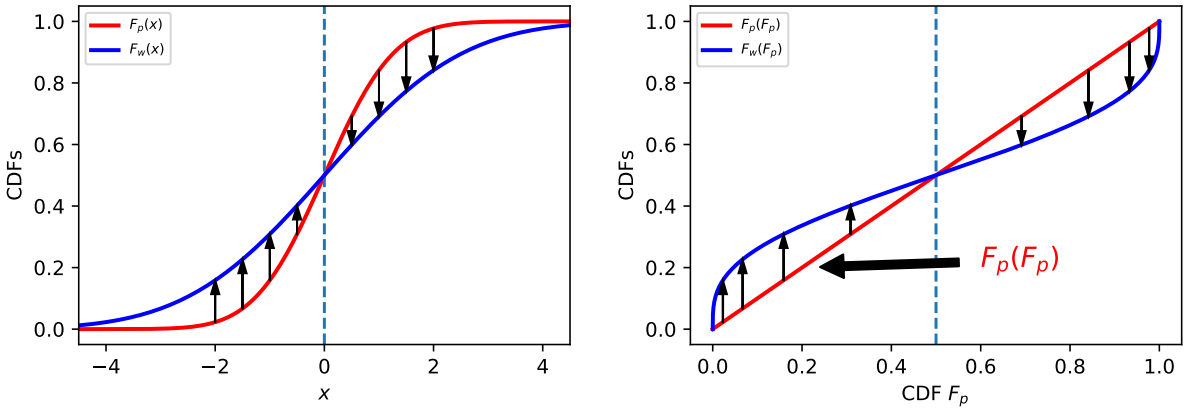

Figure 3: **Mapping CDFs.** Left: The DO models the observable as a Gaussian distributed random variable $X \sim \mathcal{N}(0,1)$, which results in the red CDF of the standard normal, $F_p(x) = \Phi_{0,1}(x)$. The DM is more cautious, in his model the same observable is modelled using a wider Gaussian distributed 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 the event probability x the DM's CDF is larger than the DO's CDF, $F_p(x) < F_w(x)$; the curves coincide at 0.5 because no difference in location is assumed; necessarily for large values of the event probability x the DM's CDF must be lower than the DO's. Right: the same CDFs as on the left but now plotted not against x but against the CDF F_p. Trivially, the CDF F_p plotted against itself is the diagonal; the CDF F_w now displays the generic inverse-S shape known from prospect theory. The arrows start and end at the same vertical values as on the left.

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, to which we will return in Sec. 4. A difference in assumed scales and locations, for 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.[5]

[5]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.

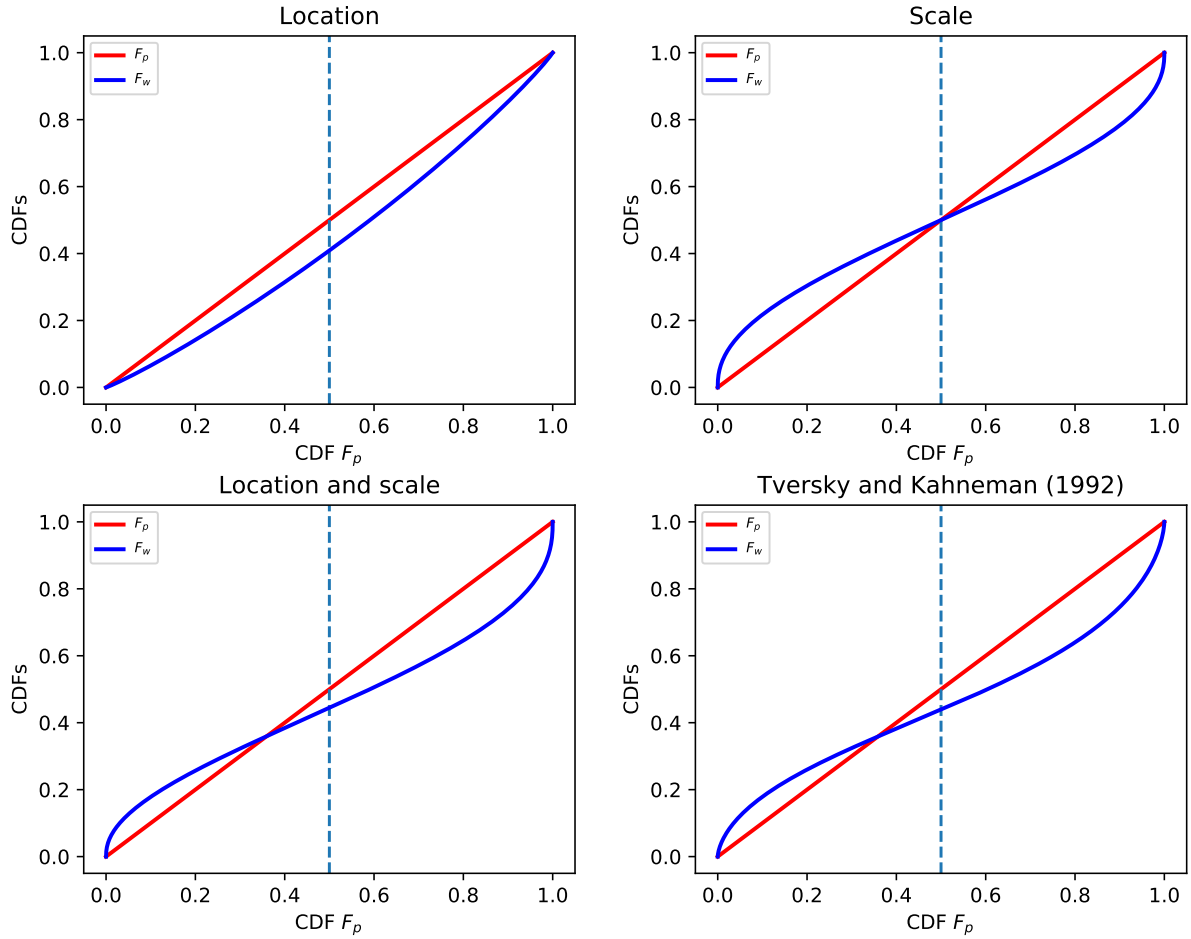

Figure 4: **CDF maps for Gaussian distributions.** Top left: Difference in location. DO assumes location 0, scale 1; DM assumes location 0.23 (bigger than DO), scale 1. Top right: Difference in scale. DO assumes location 0, scale 1; DM assumes location 0, scale 1.64 (broader than DO). Bottom left: Differences 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 Tversky and Kahneman (1992). This is (Eq. 4) with $\gamma = 0.65$. Note the similarity to bottom left.

2.3 Different shapes

Numerically, our procedure can be applied to arbitrary distributions:

1. construct a list of values for the CDF assumed by the DO, $F_p(x)$.
2. construct a list of values for the CDF assumed by the DM, $F_w(x)$.
3. plot $F_w(x)$ *vs.* $F_p(x)$.

Of course, the DM could even assume a distribution whose shape differs from that of the DO's distribution. The inverse-S arises whenever a DM assumes a greater scale for a unimodal distribution. To illustrate the generality of the procedure, in Fig. 5 we carry it out for Student's (power-law tailed) t-distributions (which we refer to as t-distributions),

where DO and DM use different shape parameters and different locations [6] The result is qualitatively similar to the bottom right panel of Fig. 3, corresponding to (Eq. 4).

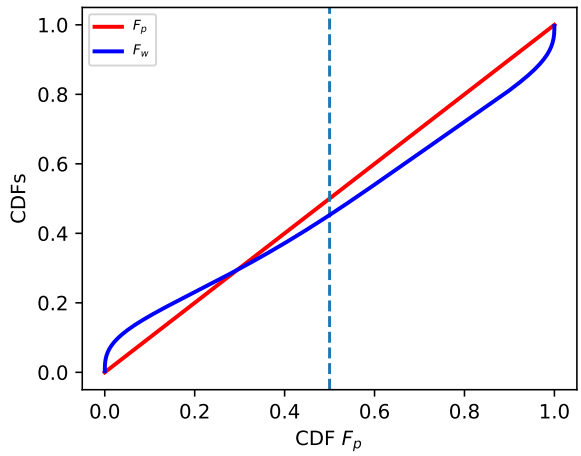

Figure 5: **Probability weighting for t-distributions.** The DM uses different shape and location parameters (1 and 0.35 respectively) from those of the DO (3 and 0.2).

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.[7] 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. Commenting on another so-called cognitive bias regarding probabilities, the representativeness fallacy, Cohen (1979) asked: "Whose is the fallacy?"

[6]The PDF of the t-distribution is

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

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 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).

[7]Indeed, its originators presented it as such. Introducing prospect theory, Kahneman and Tversky (1979, p. 277) wrote "we are compelled to assume [...] that decision weights do not coincide with stated probabilities. These departures from expected utility theory must lead to normatively unacceptable consequences". They classified prospect theory as descriptive rather than normative, *i.e.* as relating to actual rather than optimal behaviour (Tversky and Kahneman 1986, p. S252). Put simply, prospect theory aims to model systematic errors in human decision making, arising (in part) from inappropriate psychological adjustments of known probabilities.

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, and 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.

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").[8]

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 his own understanding of the rules of the game. We will return to this in Sec. 3.2.

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 an observable but must be inferred from counts.

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. How exactly it is linked to physical reality is not completely clear.

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.

[8]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).

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 of rain here 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 see 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 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 his decision making. Specifically, the DM should acknowledge that, due to his lack of information, *prima facie* rare events may be rather more common than his 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 his estimates, commensurate with his uncertainty in them. This would look like probability weighting to a DO and, indeed, would constitute a mechanistic reason for it.[9]

[9] 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.

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 A for a detailed discussion. Here we make a simple scaling argument and then check it with a simulation. For an asymptotic 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 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 probability 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 probabilities come with larger relative errors, and constitutes the key message of this section. Errors in probability estimates behave differently for low probabilities than for high probabilities: absolute errors are smaller for lower probabilities, but relative errors are larger.[10]

Let's assume that the DM is aware of the uncertainties in his estimates and, furthermore, that he does not like surprises. To avoid surprises, he adds the standard error to his estimate of the probability density, $\hat{p}(x)$, in order to construct his decision weight density, $w(x)$. In effect, he constructs a reasonable worst case for each of his 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,

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

[10] 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.

Note that the cautionary correction term in (Eq. 14) is parametrised by $T\delta x$, 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 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 reference probability density functions. Second, by simulating the DM compiling counts of outcomes drawn from reference distributions, from which he estimates probability densities and their uncertainties. The simulation is meant to explore how noisy the effect is when a DM really only sees 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 PDFs and CDF mappings generated by setting $\hat{p}(x)$ in (Eq. 14) to be the probability density functions for 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 he 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, $\varepsilon\,[\hat{p}(x)]$, is obtained numerically as standard deviation in each $\hat{p}(x)$ over 1000 parallel simulations. The DM's decision weights are then obtained according to (Eq. 13). Again, inverse-S curves are found for both distributions, corroborating our scaling arguments.

3.2.1 Typical situations of DO and DM: ergodicity

To recap: behavioural economists observe that DOs tend to assign lower weights to low-probability events than DMs. While behavioural economists commonly assume 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 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, he is concerned with what happens to him 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 stochastic process that is not ergodic. The ensemble average of wealth does not behave like the time average of wealth.

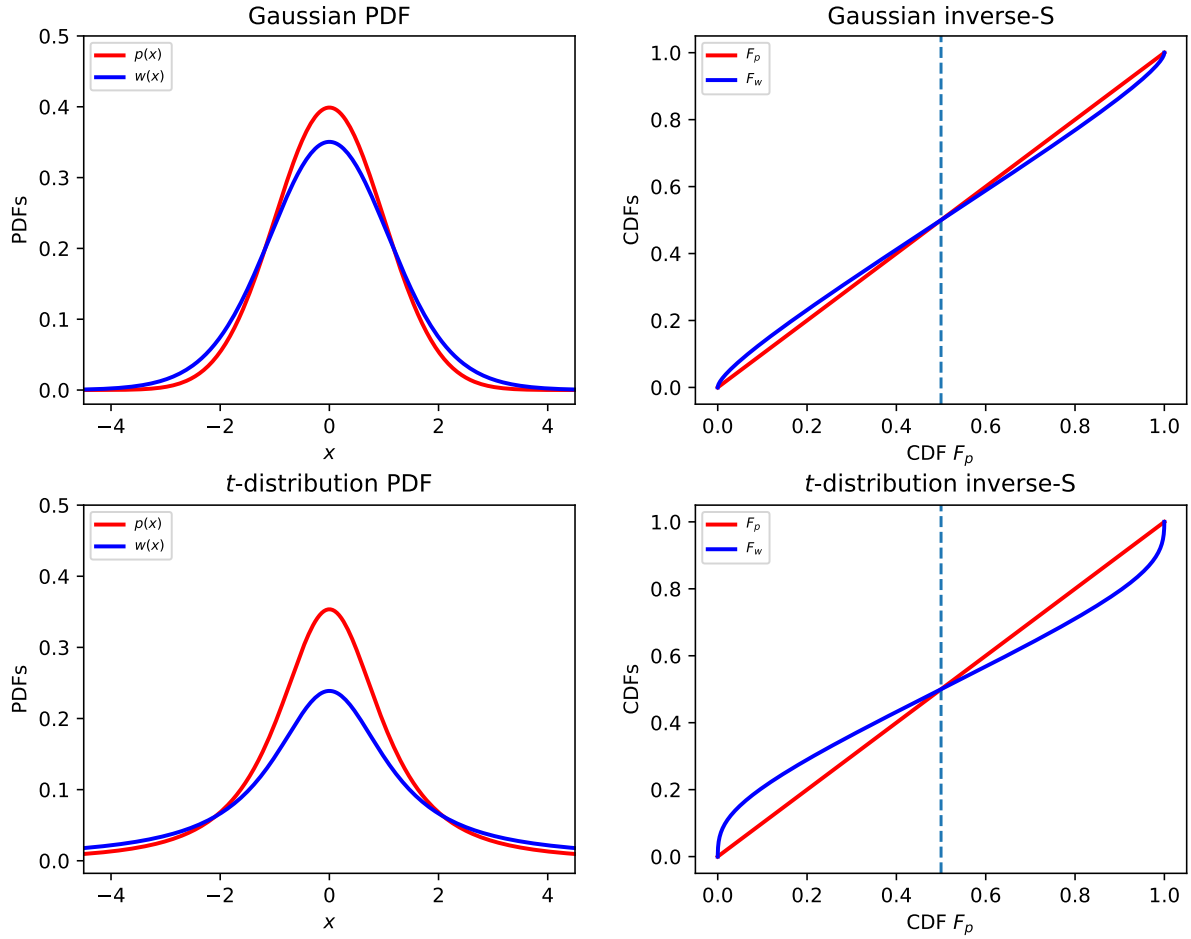

Figure 6: **Mapping PDFs and CDFs with estimation errors.** PDFs (left) and inverse-S curves (right) arising when the DO assumes a Gaussian (scale 1, location 0, top line) or a t-distribution (shape 2, location 0, bottom line), and the DM uses decision weights according to (Eq. 14) with $T\delta x =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

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 low-probability 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.

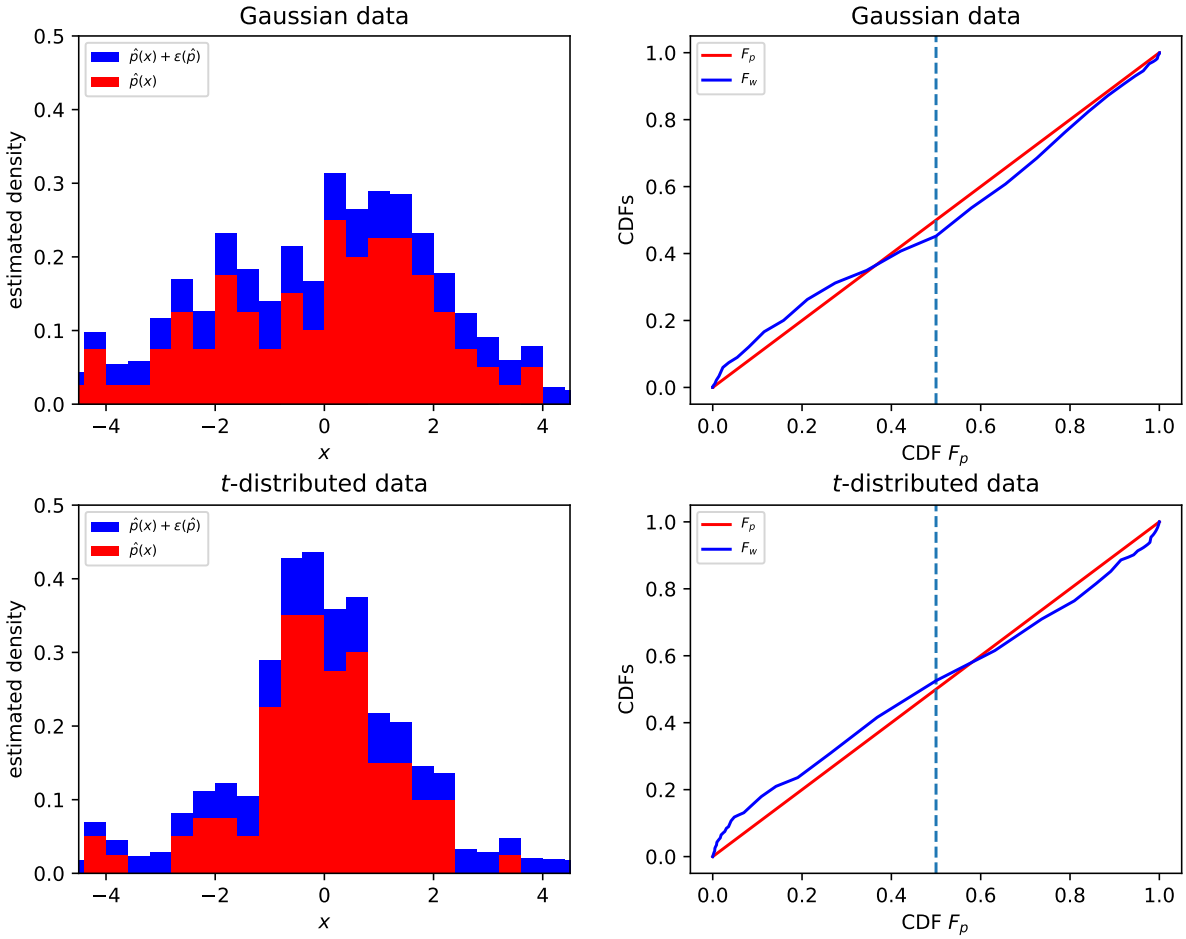

Figure 7: **Simulations of a DM estimating probability densities by counting events in a finite series of observations.** Left: estimated probability densities for $T = 100$ Gaussian (top; location 0, scale 2) and t-distributed (bottom; location 0, scale 1, shape 1.5) variates counted in bins of width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

4 Fitting the model to experimental results

Visually, looking at the figures and the level of noise in the data in Fig. 1, one would conclude that Tversky and Kahneman's physically unmotivated function, $F_w^{TK}(F_p)$ in (Eq. 4), fits the data no more efficiently than the functions arising from our mechanistic model. This is particularly evident in the bottom panels of Fig. 4, which show that a Gaussian, $w(x)$, whose scale and location differ from those of $p(x)$, reproduces the fitted functional shape of $F_w^{TK}(F_p)$.

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 Tversky and Kahneman (1992) (depicted in circles 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 (Eq. 4) used by Tversky and Kahneman, 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 parametrically describe probability weighting (widely used also *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 models are characterised by two parameters, whereas (Eq. 4) only has one free parameter. Equation (16) has two parameters.

Figure 8 presents the fit results. We obtain very good fits to data for both Gaussian and t-distributions, as well as for (Eq. 4) and (Eq. 16), in the two experiments. 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 simply phenomenological.

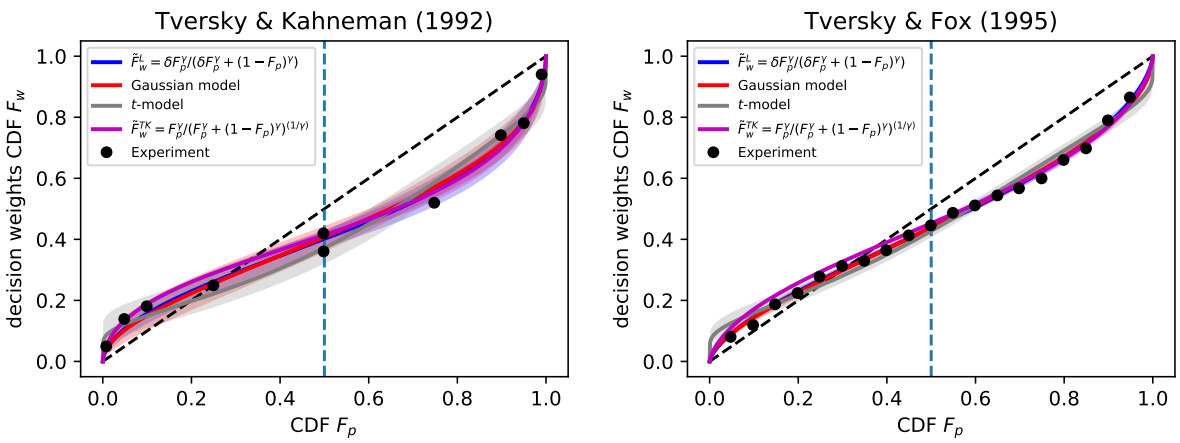

Figure 8: **Model fitting to experimental data from Tversky and Kahneman (1992) (left) and Tversky and Fox (1995) (right).** Left: Lattimore et al. (1992) (Eq. 16): $\delta = 0.67$ ($SE = 0.04$), $\gamma = 0.58$ (± 0.03); Gaussian model: $\mu = 0.38$ (± 0.06), $\sigma = 1.60$ (± 0.10); t-model: $\nu = 1.27$ (± 0.28), $\mu = 0.40$ (± 0.07); Tversky and Kahneman (1992) (Eq. 4): $\gamma = 0.60$ (± 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 non-linear least squares curve fitting.

5 Discussion

On 28 February 2020, Sunstein (2020), a behavioural economist, legal scholar, and former United States Administrator of the Office of Information and Regulatory Affairs, diagnosed that people's concern about a potential coronavirus outbreak in the US was attributable to an extreme case of probability weighting. Supposedly, according to Sunstein, people were neglecting the fact that such an event had a low probability. When the piece was published, many commented that it seemed quite reasonable to them to take precautions, and that Sunstein himself may have underestimated both the severity and likelihood of what lay ahead. One month later, the US suffered a major outbreak of coronavirus.

This sad episode illustrates that an inverse-S curve is a neutral indicator of a difference in opinion. It says nothing about who is right and who is wrong.

The term "probability weighting" suggests an obscure mental process, where a DM carries out operations on probabilities. It seems more natural to us to consider a DM modelling events about whose probabilities he 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 the 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" is purely descriptive and comes with the ill-conceived connotation of DMs suffering from a cognitive error. The phenomenon is better thought of 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 in time.

References

Adamou, Alexander, Yonatan Berman, Diomides Mavroyiannis, and Ole Peters (2020). "Microfoundations of Discounting". *arXiv:1910.02137*. arXiv: 1910.02137 (cit. on p. 5).

Berman, Yonatan and Mark Kirstein (2020). "Risk Preferences in Time Lotteries". *researchers.one*. URL: bit.ly/TimeLotteries (cit. on p. 5).

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. 4).

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 p. 10).

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. 4).

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. 4).

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. 11).

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. 11).

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. 11).

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. 4, 13).

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 p. 4).

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, 4, 10, 12).

Laplace, Pierre Simon (1814). *Théorie analytique des probabilités*. 2nd ed. Paris, Ve. Courcier (cit. on p. 4).

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 p. 17).

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. 17).

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. 4).

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. 5).

Von Neumann, John and Oskar Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press (cit. on p. 4).

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. 5).

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. 5).

Peters, Ole (2019). "The Ergodicity Problem in Economics". *Nature Physics* 15 (12), 1216–1221. DOI: 10.1038/s41567-019-0732-0 (cit. on pp. 5, 14).

Prelec, Dražen (1998). "The Probability Weighting Function". *Econometrica* 66 (3), 497–527. DOI: 10.2307/2998573 (cit. on p. 17).

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 p. 4).

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. 8).

Savage, Leonard J. (1954). *The Foundations of Statistics*. Dover Publications, New York (cit. on p. 4).

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. 4).

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. 4).

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. 18).

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. 16, 17).

Tversky , Amos and Daniel Kahneman (1986). "Rational Choice and the Framing of Decisions". *Journal of Business* 59.4, S251–S278. JSTOR: 2352759 (cit. on p. 10).

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. 2–5, 9, 16, 17).

Tversky , Amos and Peter Wakker (1995). "Risk Attitudes and Decision Weights". *Econometrica* 63 (6), 1255–1280. DOI: 10.2307/2171769 (cit. on p. 17).

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. 4).

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. 4).

A 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's connect this statement to one about relative uncertainties. The decision weight is arrived at by adding the probability $p(x)$ to its uncertainty $\varepsilon\left[p(x)\right]$ and normalising, as we did in (Eq. 13), *i.e.*

$$w(x) = \frac{p(x) + \varepsilon\left[p(x)\right]}{\int_{-\infty}^{\infty}\left(p(s) + \varepsilon\left[p(s)\right]\right) ds} \,. \tag{17}$$

This can be expressed as

$$w(x) = p(x)\left(\frac{1 + \frac{\varepsilon[p(x)]}{p(x)}}{\int_{-\infty}^{\infty} p(s)\left\{1 + \frac{\varepsilon[p(s)]}{p(s)}\right\} ds}\right) \,, \tag{18}$$

where $\frac{\varepsilon[p(x)]}{p(x)}$ is the relative error, and the denominator of (Eq. 18) is a normalisation constant. If the relative error is larger for small probabilities than for large probabilities, then small probabilities are enhanced more (the summand $\frac{\varepsilon[p(x)]}{p(x)}$ in the numerator is greater) than large probabilities. The normalisation constant scales down all probabilities equally, and where the enhancement was greater, $w(x)$ ends up above $p(x)$, and where it was lower $w(x)$ ends up below $p(x)$. So, if the relative error is larger for small probabilities, 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. 18) as

$$w(x) = p(x)\left(\frac{1 + \frac{\varepsilon[p]}{p}}{1 + \left\langle\frac{\varepsilon[p]}{p}\right\rangle}\right) \,, \tag{19}$$

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

B How to draw an 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 his model of uncertainty than the DO. In the Gaussian case, it was possible to express the decision weight w as a function of the probability p,

$$w(p) = p^{\frac{1}{\alpha^2}} \frac{(2\pi\sigma^2)^{\frac{1-\alpha^2}{2\alpha^2}}}{\alpha} \ , \tag{20}$$

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. 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 $\alpha > 1$). This situation would arise if the DM were an "insider" with better information about the decision problem than provided to him by the DO. Indeed, the prediction of an S curve under such conditions is an empirically testable prediction, which distinguishes our mechanism for probability weighting from its classical presentation.

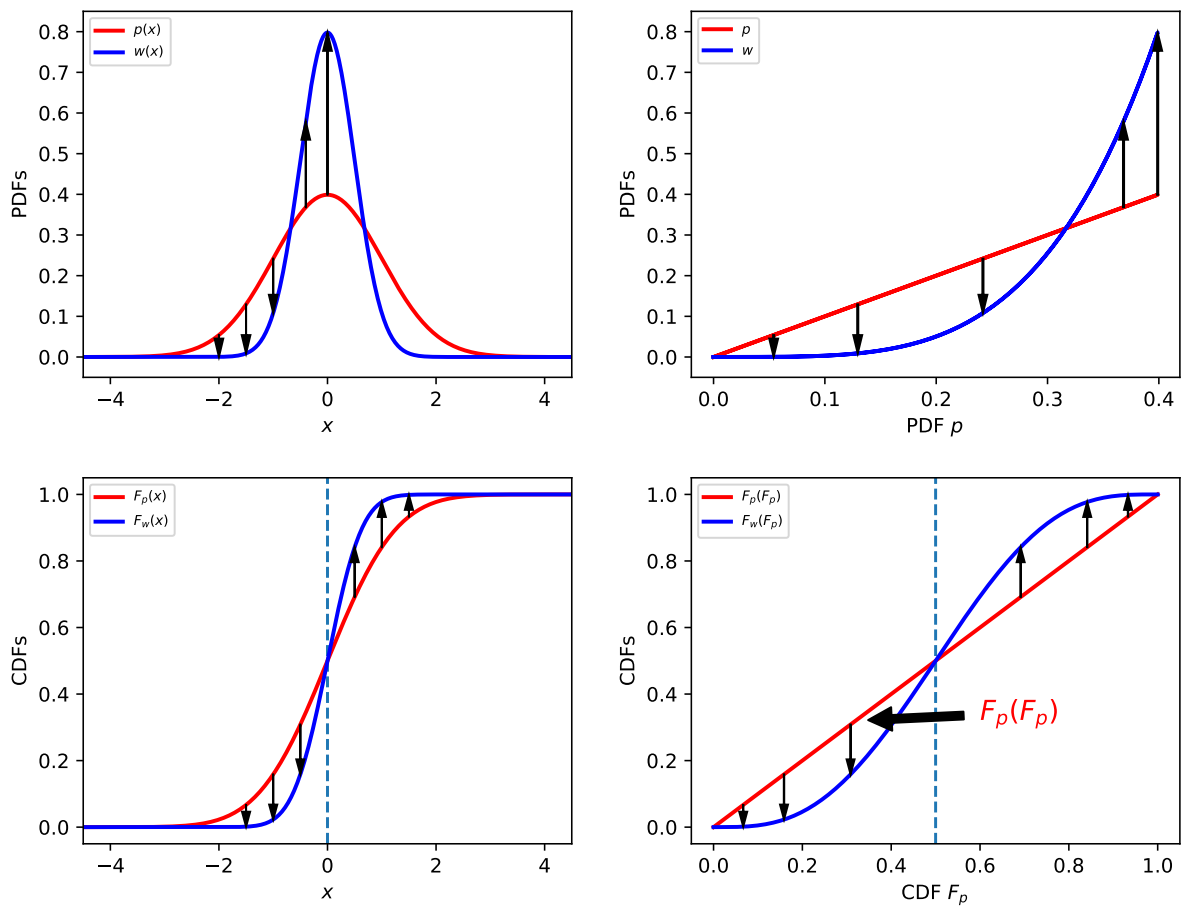

Figure 9: **Generating an S curve.** Analogues of Fig. 2 and Fig. 3 for the case $0 < \alpha < 1$.