What are we weighting for? A mechanistic model for probability weighting

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

April 13, 2020

Abstract

Behavioural economics collects observations of human economic behaviour and provides labels for those observations. Probability weighting is one such label. It expresses a mismatch between probabilities used in a formal model of a decision problem (*i.e.* model parameters) and probabilities inferred from real people's behaviour faced with the modelled decision problem (the same parameters estimated empirically). The inferred probabilities are called "decision weights." It is considered a robust observation that decision weights are higher than probabilities for extreme events, and (necessarily, because of normalisation) lower than probabilities for common events. The observed behaviour thus amounts to the refusal by real decision-makers totally to rely on a formal model, and instead to exercise extra caution. In this paper we explore quantitatively how such caution generically reproduces existing empirical findings. We list quantitatively well-specified reasons for such caution and find the resulting probability weighting, as a benchmark for reasonable behaviour.

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

JEL Codes C61 · D01 · D81

*London Mathematical Laboratory, 8 Margravine Gardens, London W6 8RH, UK and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, 87501 NM, USA. Email: o.peters@lml.org.uk

†London Mathematical Laboratory, 8 Margravine Gardens, London W6 8RH, UK. Email: a.adamou@lml.org.uk

‡London Mathematical Laboratory, 8 Margravine Gardens, London W6 8RH, UK. Email: m.kirstein@lml.org.uk

§London Mathematical Laboratory, 8 Margravine Gardens, London W6 8RH, UK. Email: y.berman@lml.org.uk

1 Nomenclature

Probability weighting is a concept that originated in prospect theory. It is one way to conceptualize a pattern in human behavior, of caution with respect to formal models. This is best explained with an example:

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

that an event occurs with some probability. The DM's behaviour is then observed, and is found to be consistent with a behavioural model (for example expected-utility optimization) where the DM uses a probability that differs systematically from what the DO has declared. Specifically, it is consistently observed that DMs act as though extreme events (those of low probability) had higher probabilities than what's specified by the DO. These apparent "higher probabilities" are called "*decision weights*" because they are better at describing the decisions actually made than the probabilities specified by the DO. 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.[1]

The key observation of $w(x) > p(x)$ for small p *etc.* is often summarised visually with a comparison between

- 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).

As a final piece of nomenclature, we will use the terms *location*, *scale*, and *shape* when discussing probability distributions. Consider a standard normal distribution $\mathcal{N}(0,1)$ – here, the 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 parameters for location μ_X and scale σ_X, the transformation in (Eq. 3) obtains

[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).

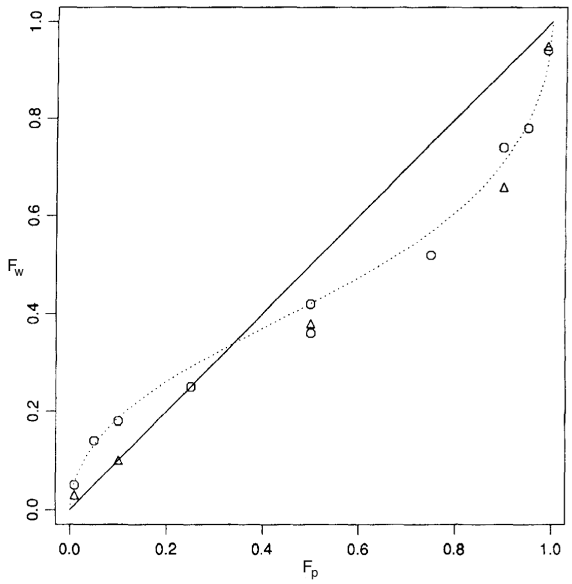

Figure 1: **Empirical phenomenon of probability weighting.** Cumulative decision weights F_w (used by decision makers) versus cumulative probabilities F_p (used by disinterested observers), as reported by (Tversky and Kahneman 1992, p. 310, Fig. 1). The figure is to be read as follows: pick a point along the horizontal axis (the cumulative probability used by a DO) and look up the corresponding value on the vertical axis of the dotted inverse-S curve (the cumulative decision weight 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.

the identically-shaped location-0 and scale-1 distribution for the so standardised random variable

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

Thus the PDF of Z, $p(z)$ is a density 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.

2 Consistent probability weighting as a difference between models

Behavioural economics interprets Fig. 1 as evidence for a cognitive bias of the DM. We will keep a neutral stance. We don't assume the DO to know "the truth" – he has a model of the world. Nor do we assume the DM to know "the truth" – he has another model of the world. 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[2]

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

We note that no mechanistic motivation was given for fitting this specific family of CDFs (parameterised by γ). The motivation is purely phenomenological: with $\gamma < 1$, this family "looks a bit like the data." The function $F_w^*(F_p;\gamma)$ has only one free parameter, γ. For $\gamma = 1$ it is the identity, and the CDFs coincide, $F_w^*(F_p) = F_p$. The function F_w^* 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. Because of this, 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. We will see below that the curvature and shift correspond to different parts of a mechanistic explanation of the phenomenon.

2.1.2 Mechanistic explanation of the inverse S

Specifically, we now make explicit how the robust qualitative observation of the inverse-S shape in Fig. 1 emerges from assuming that the DM uses a larger scale in his model of the world than the DO. This can have numerous reasons, to which we will return in Sec. 3. For now, suffice it to say that precaution is an obvious one: any uncertainty the DM wishes to include in his model in addition to what the DO includes will translate into a greater scale for the DM's distribution and therefore into an inverse-S shape for any unimodal (peaked) distribution when cumulative densities are compared.

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 with location μ and variance σ^2. And let's further assume that a DM (for whatever reason, perhaps caution) models the same observable as a Gaussian with the same location, μ, but with a greater scale, so that the variance is $\alpha^2\sigma^2$. The DM simply assumes a broader range – α times greater – of plausible values, left panel of Fig. 2.

Generically, if the DM is using a greater scale in his model, then he is using higher decision weights for low-probability events, and (because of normalisation), lower decision weights for high-probability events than the corresponding model of the DO. 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

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

and

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

[2]Equation (4) is the consensus functional form in the community Barberis (2013). ****OP: Maybe not quite, given that we fit another function later. We could introduce Latimore's function hiere. OP****

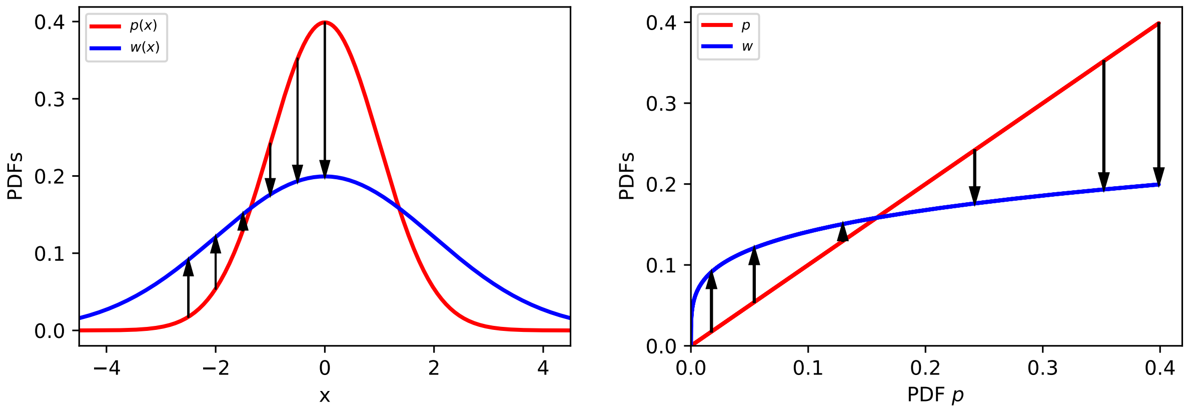

Figure 2: **Mapping PDFs.** Left: probability PDF (red), estimated by a DO; and decision-weight PDF (blue), estimated by a DM. The DO models x with a best estimate for the scale (standard deviation) and assumes the true frequency distribution is the red line. The DM models x with a greater scale (here 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 assigned probabilities can also be expressed by directly plotting, for any value of x, the two different PDFs against one another. 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.

solve (Eq. 6) for $(x-\mu)^2$, substitute that in in (Eq. 5), and obtain the following expression for decision weights directly as a function of probabilities

$$w(p) = p^{\frac{1}{\alpha^2}} \frac{(2\pi\sigma_1^2)^{\frac{1-\alpha^2}{2\alpha^2}}}{\alpha} , \tag{7}$$

which is precisely what's plotted 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 arguments and below it for large arguments.

Alternatively, we can express the difference between models by plotting the CDFs F_w and F_p. We do this in Fig. 3, revealing the origin of the inverted S as the DM assuming a greater scale.

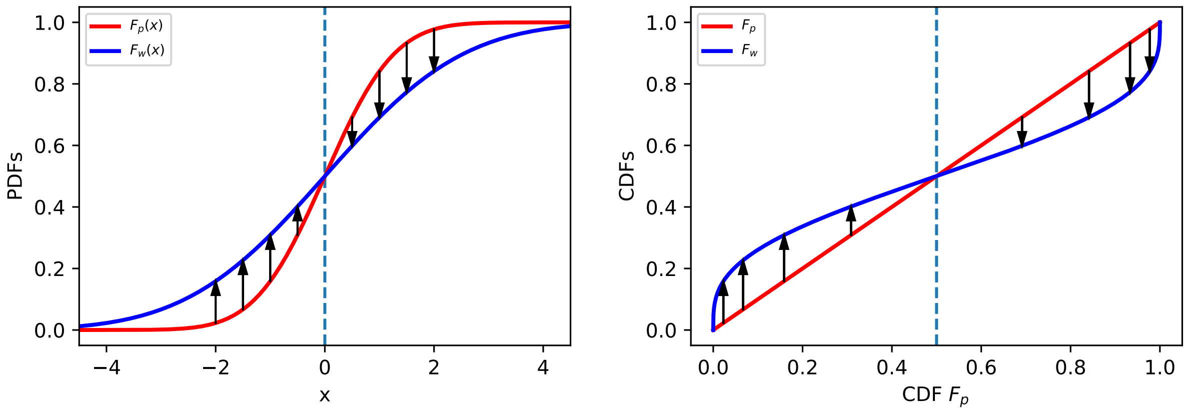

Figure 3: **Mapping CDFs.** Left: The DO assumes the observable X follows Gaussian distribution $X \sim \mathcal{N}(0, 1)$, which results in the red CDF of the standard normal, $F_p(x) = \Phi_{0,1}(x)$. The DM is more cautious, in his model the same observable X follows a wider Gaussian distribution, $X \sim \mathcal{N}(0, 3)$ depicted by F_w (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. Because the horizontal axis is has been non-linearly stretched (as the argument changed from x to F_p), their horizontal locations are shifted.

2.2 A mismatch between both 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.

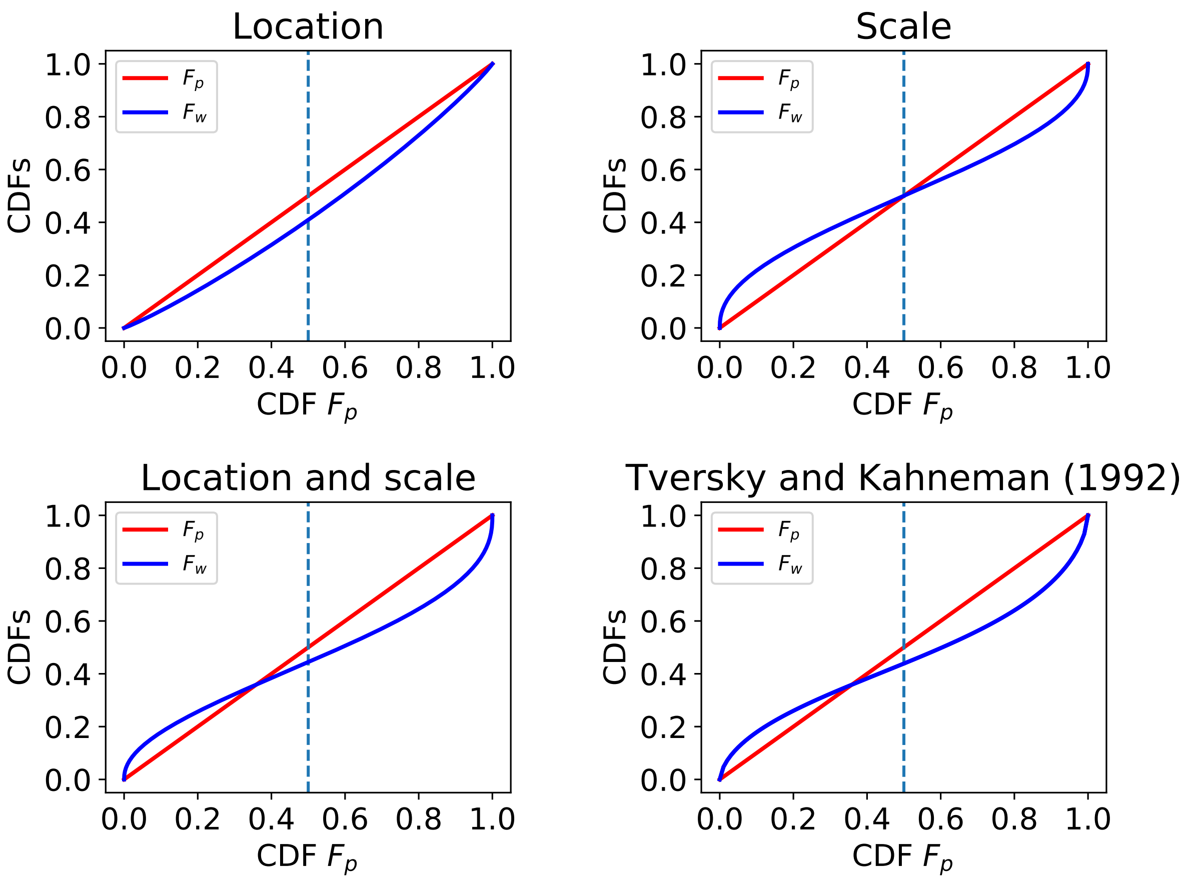

Figure 4: **Decision weight CDFs used by a DM vs. probability CDFs used by a DO, Gaussian distribution**.
Top left: Difference in scale. DO assumes location 0, scale 1; DM assumes location 0, scale 1.64 (broader than DO).
Top right: Difference in location. DO assumes location 0, scale 1; DM assumes location 0.18 (bigger than DO), scale 1.
Bottom left: Differences in scale and location. DO assumes location 0, scale 1; DM assumes location 0.18 (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, and fat-tailed distributions

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. An infinity of combinations of assumed distributions can be explored. 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 a (power-law tailed) Student's t-distribution, where DO and DM use different shape parameters and different locations.[3] The result is qualitatively similar to the bottom right panel of Fig. 3, corresponding to (Eq. 4).

A difference in assumed scales and locations 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 just means that experimenters and their subjects have different views about what might be an appropriate response to a situation.

On 28 February 2020, Cass Sunstein (2020), a behavioral economist 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 – they neglected the fact, supposedly, 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 Sunstein himself may have underestimated the severity of what lay ahead. One month later the US became the epicentre of the global coronavirus pandemic. The episode illustrates that an inverted S-curve is a neutral indicator of a difference in opinion. It says nothing about who is right and who is wrong.

[3]The pdf of the Student's t-distribution is

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

where ν is the shape parameter and μ is the location parameter. The corresponding cdf is

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

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

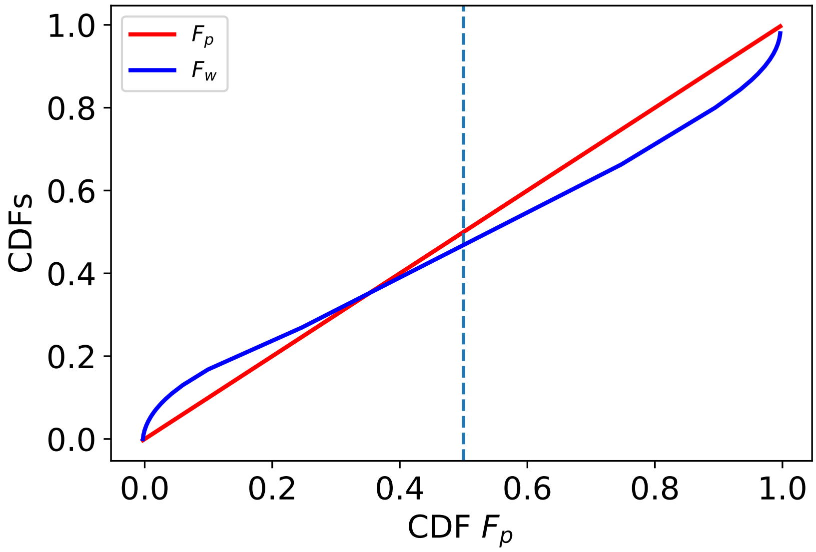

Figure 5: Probability weighting for Student-t distributions, where the DM uses a different shape parameter (1) and a different location parameter (0) from those of the DO (2 and 0.2, respectively).

3 Reasons for different models

"Probability weighting" as a term is suggestive of a detrimental cognitive bias. We caution against this interpretation. At the very least it must be borne in mind that it is unclear who suffers from the bias: experimenters or test subjects? It is interesting, however, that the difference in opinion tends to go in the same direction. In more neutral language, DMs tend to assume a greater range of plausible outcomes than DOs. Why might this be? Two conditions must be satisfied. First, there must be a reason for frequent disagreement about probabilities used in a model; second, there must be a reason for such disagreement to be consistent: there must be a relevant systematic difference between the DO and the DM.

The first condition is satisfied because the word "probability" is commonly interpreted in different ways. Even once one has settled on a definition, numerical values for probabilities are still difficult to estimate from real-world observations (see Sec. 3.1).

The second condition is satisfied as follows. A DO has to build a formal model, and will include a number of sources of uncertainty in it, but often not all sources. A DM instead has to make decisions in the real world, and rules of thumb like the precautionary principle (better err on the side of caution) will often lead to additional assumed uncertainty. For example, the DO may know the true probabilities of some gamble in an experiment; the DM may in addition have doubts about the DO's sincerity, or about his (the DM's) understanding of the rules of the game. We will return to this in Sec. 3.2.

3.1 "Probability" can mean different things

Many thousands of pages have been written about the meaning of probability. We will not attempt a summary of the philosophical debate and instead focus on a few relevant points.

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 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 "70% chance of rain tomorrow" refers to the relative frequency over time. Before the advent of computer models in weather forecasting, people used to compare recent measurements (of wind and pressure today, say) to measurements further in the past – weeks, months, years earlier, that were similar and where one had reason to believe that what had happened 1 day later would be similar to what will happen tomorrow.

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 90% sure I left my wallet in that taxi."

For our purpose it suffices to say that there's no guarantee that a probabilistic statement will be interpreted by the listener (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

Imagine the DO and DM have the same interpretation of the word "probability." Say they agree that they mean 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 never known with certainty. They can only be estimated, by counting. If an event occurs in n out of T observations, then our best estimate for its probability is n/T

As the probability of an event gets smaller, so does the number of times we see it in a finite time series. Moreover, fluctuations in this number – and, therefore, the uncertainty in our estimate of the probability – become relatively larger as an event becomes rarer. Take an extreme example of an event which occurs in 0.1% observations asymptotically, of which we have a time series of 100 observations. Almost all such series (around 99.5% of them) will contain 0 or 1 events, meaning that we will estimate the probability as either 0 or 1%. In other words, we estimate the event as either impossible or occurring ten times more frequently than it really would in a long series. If, however, the event occurs 50% of the time asymptotically, then a probability estimate from 100 observations would likely (in around 95% of series) be in the range 40–60%, leading to a much smaller relative error.

If the relative uncertainty in a probability estimate becomes a larger as an event becomes rarer, then a DM who must estimate probabilities from observations is well advised to account for this in his decision-making. Specifically, he should acknowledge that, due to his lack of information, *prima facie* rare events may be rather more common than his data imply, while common events, being revealed more often, are more easily characterised. In such circumstances, caution may dictate that he assign higher probabilities to rare events than his estimates, commensurate with his uncertainty in those estimates. This would look like probability weighting and, indeed, would constitute a mechanistic reason for it.

Let's build a simple model to see if it matches the stylised facts of probability weighting, including the inverse-S curve. We model the arrival of events as a Poisson process, whose rate parameter is the asymptotic probability, p, of the event. Put simply, this means that the count per unit time converges to p over long time. The count, $n(T)$, over the finite observation time, T, is a Poisson-distributed random variable with intensity, pT. It has mean value, pT, and standard deviation, $\sqrt{pT}$. If the DM makes one observation per time unit, then the measured relative frequency, $\hat{p} = n(T)/T$, is an unbiased estimator of the asymptotic probability, p, with standard error, $\sqrt{\hat{p}/T}$.

The standard error in an estimated probability shrinks as the probability decreases. However, the relative error in the estimate is $1/\sqrt{\hat{p}T}$, which grows as the event becomes rarer. This is consistent with our claim, that low probabilities come with larger relative errors, further exemplified in Tab. 1.

Asymptotic probability	Most likely count	Standard error in count	Standard error in probability	Relative error in probability
0.1	1000	32	0.003	3%
0.01	100	10	0.001	10%
0.001	10	3	0.0003	30%
0.0001	1	1	0.0001	100%

Table 1: This table assumes $T = 10000$ observed time intervals. To be read as follows (first line): for an event of asymptotic probability 0.1, the most likely count in 10000 trials is 1000. Assuming Poisson statistics, this comes with an estimation error of $\sqrt{1000} = 32$ in the count and $32/10000 = 0.003$ in the probability, which is $0.003/0.1 = 3\%$ of the asymptotic probability.

The most important message is that 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.

Let's make one more assumption: DMs don't like surprises. To avoid surprises, a DM will incorporate the uncertainty in probability estimates by assuming that the actual long-time relative frequency of an event is his best estimate plus two standard errors. In the notation of density functions, suppose he estimates the probability of an event, $x \leq X \leq x + \delta x$, as $\hat{p}(x)\delta x$ from a time series of length T. To this he adds two standard errors to get his decision weight for the event,

$$w(x)\delta x = \hat{p}(x)\delta x + 2\sqrt{\frac{\hat{p}(x)\delta x}{T}} \simeq \left(\hat{p}(x) + \sqrt{\frac{\hat{p}(x)}{T}}\right)\delta x \qquad (10)$$

for small bin width, $\delta x \ll 1$. Repeating this procedure across the domain of the density function, *i.e.* for all possible events, gives a decision weight density of

$$w(x) = \frac{\hat{p}(x) + \sqrt{\hat{p}(x)/T}}{\int_{-\infty}^{\infty} [\hat{p}(s) + \sqrt{\hat{p}(s)/T}] ds}, \tag{11}$$

where the denominator ensures normalization. Note the role of T here: as T grows large, *i.e.* as the DM collects more information, the cautionary correction term vanishes, and both $w(x)$ and $\hat{p}(x)$ converge to the asymptotic density, $p(x)$. With perfect information, a DM does not need to adjust decisions to account for uncertainty.

Figure 6 shows the resulting PDFs and the CDF mappings, for a Gaussian distribution and for a fat-tailed Student-t distribution. As one might expect, the effect is more pronounced for the fat-tailed distribution.

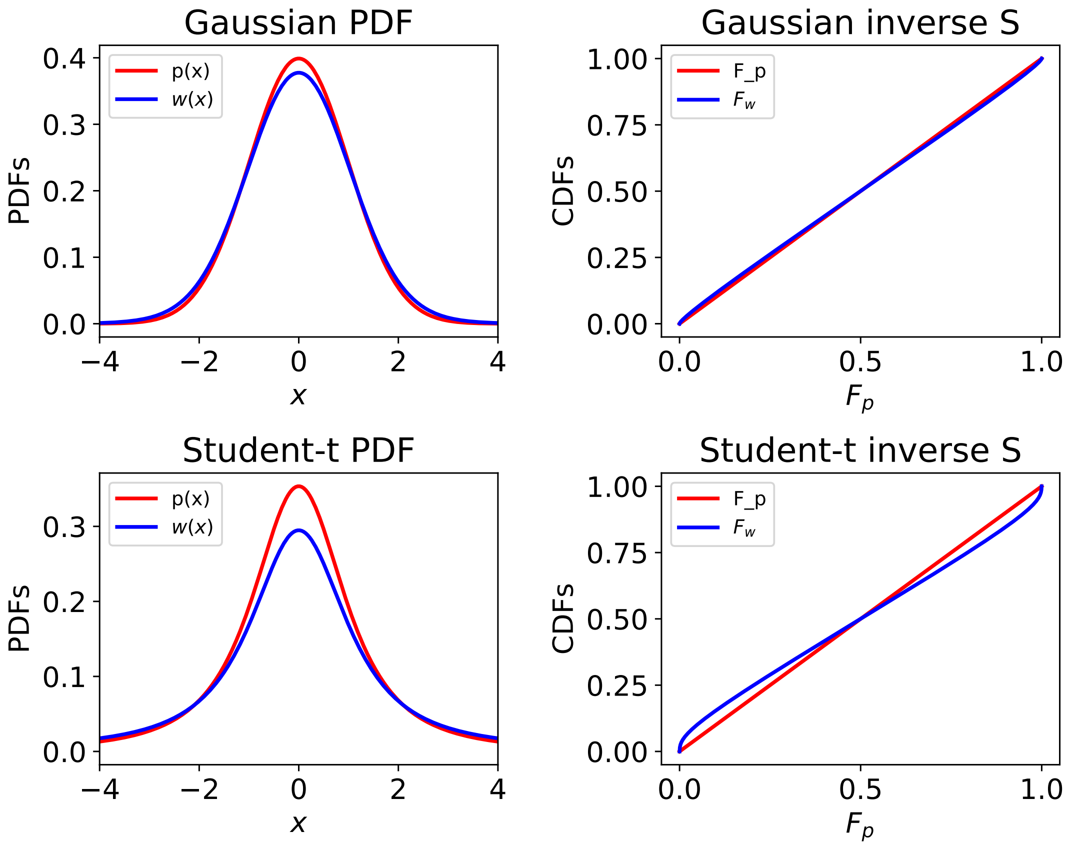

Figure 6: PDFs and inverse-S curves arising when the DO assumes a Gaussian (scale 1, location 0, top line) or a Student-t distribution (shape 2, location 0, bottom line), and the DM uses decision weights according to (Eq. 11) with $T = 100$ observations. For the fat-tailed Student-t distribution the difference between p and w is more pronounced.

3.2.1 Typical situations of DO and DM: ergodicity

To recap: behavioral economics observes that DOs tend to assign lower weights to low-probability events than DMs. While behavioral economics commonly assumes that the DM is wrong, we make no such judgement. In any decision problem, the aim of the decision must be taken into account, and that crucially depends on the situation of the individual.

The two types of modellers (DO and DM) pursue different goals. Broadly, the DO tends to be a behavioral scientist without personal exposure to the success or failure of the DM (who tends to be a test subject or someone whose behavior 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 usually not a primary concern – instead, the concern of the DM is what happens to him over time. Not distinguishing between these two perspectives is only permissible if they lead to identical predictions, and that is only the case in ergodic situations (Peters 2019).

It is now well known that the situation usually studied in decision theory is not ergodic 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.

The most striking example is the universally important case of noisy multiplicative growth – universal because it is the fundamental process that drives evolution: noise generates the diversity of phenotypes necessary for evolution and multiplicative growth (self-reproduction) is how successful phenotypes spread their traits in a population. This process operates on amoeba, as it does on forms of institutions, and on investment strategies.

The simplest model of noisy multiplicative growth is geometric Brownian motion, $dx = x(\mu dt + \sigma dW)$. The average over the full statistical ensemble (often studied by the DO) of geometric Brownian motion grows as $\exp(\mu t)$. The individual trajectory of geometric Brownian motion, on the other hand, grows in the long run as $\exp[(\mu - \frac{\sigma^2}{2})t]$.

In the DO's ensemble perspective, noise does not affect growth and is often deemed irrelevant. In the DM's time perspective, noise reduces growth. While a DO interested in the ensemble may get away with disregarding low-probability events, for the DM's success hedging against them is of crucial importance.

The difference between how these two perspectives evaluate the effects of noise (*i.e.* of the 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 low-probability (extreme) events had a higher probability than in the DO's model.

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 fit of F_w^*, (Eq. 4), resembles the data no better than our mechanistically constrained model. This is particularly evident in the bottom two 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^*.

For completeness and scientific hygiene, in the present section we fit location and scale parameters in the Gaussian and Student's-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

$$F_w\left(F_p\right) = \Phi\left(\frac{\Phi^{-1}\left(F_p\right) - \mu}{\sigma}\right), \tag{12}$$

where Φ is the CDF of the standard normal distribution. In the Student's-t model we fit the location parameter μ and the shape parameter ν in the CDF $F_w(x)$ of a Student's t-distributed random variable (see Sec. 2.3), assuming p follows a standard normal distribution.

In addition, we fit (Eq. 4), used by Tversky and Kahneman, and the function

$$F_w^* = \frac{\delta F_p^\gamma}{\delta F_p^\gamma + (1 - F_p)^\gamma}, \tag{13}$$

suggested by Lattimore et al. (1992) to parametrically describe probability weighting, and was commonly used since (see, *e.g.* Tversky and Wakker (1995)). We fit (Eq. 13) and not only (Eq. 4), to allow a fair comparison between the Gaussian and Student's-t models, characterised by two parameters.

Figure 7 presents the fit results. We obtain very good fits to data for both Gaussian and Student's-t distributions, as well as for (Eq. 4) and (Eq. 13), 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. 13), the fitted functions are directly derived from a physically plausible mechanism, and are not simply phenomenological.

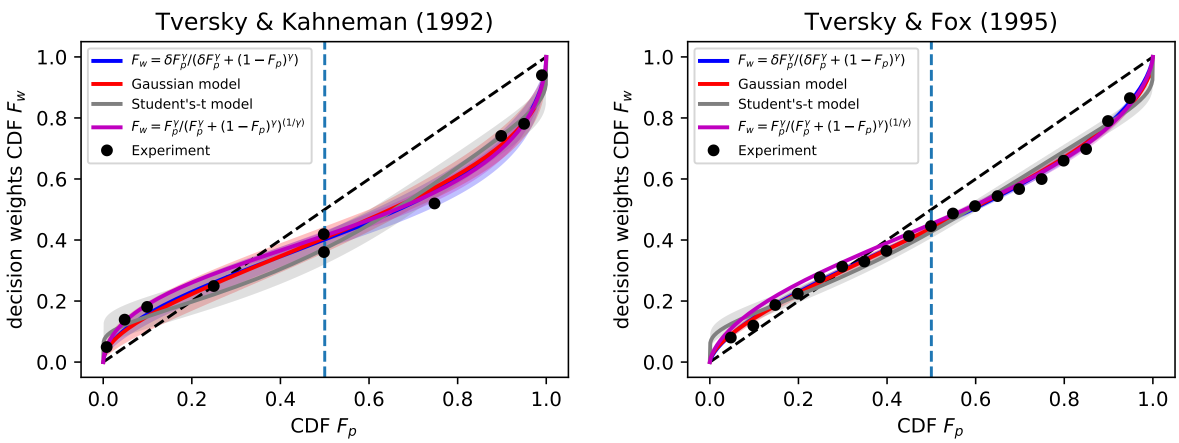

Figure 7: Model fitting to experimental data from Tversky and Kahneman (1992) (left) and Tversky and Fox (1995) (right). Left) Lattimore et al. (1992): $\delta = 0.67$ ($SE = 0.04$), $\gamma = 0.58$ (± 0.03); Gaussian model: $\mu = 0.38$ (± 0.06), $\sigma = 1.60$ (± 0.10); Student's-t model: $\nu = 1.27$ (± 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); Student's-t model: $\nu = 1.41$ (± 0.21), $\mu = 0.22$ (± 0.03); Tversky and Kahneman (1992) (Eq. 4): $\gamma = 0.68$ (± 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 Conclusion/Summary

Related literature Prelec (1998); Wu and Gonzalez (1996): derive PW from preference axioms, goal is to generate concavity and then convexity

Stott (2006): test several functional forms, corroborate Prelec (1998) and (Eq. 4) and (Eq. 13), respectively. None of the tested functional forms corresponds to the function we derived mechanistically in (Eq. 7).

Wakker (2010): the whole book contains only parameter calibrations of (Eq. 4) but provides no mechanistic explanation.

Abdellaoui et al. (2011): maybe closest to what we do in terms of their Fig. 2., however they distinguish between source functions for ambigious probabilities and known probabilities. They interpret the shifts and shape of the inverse-S curve using two additional idiosyncratic psychological parameters encoding optimism(pessimism) and likelihood (in)sensitivity.

Results set in perspective The consensus view hold in behavioural economics is that a descriptive decision model needs to feature the so-called fourfold pattern of risk attitudes:

1. gain/loss asymmetry (kink at the reference point),
2. nonlinearity in probability weighting (inverse-S curve),
3. risk aversion for most gains & low probability losses (concavity) and
4. risk seeking for most losses & low probability gains (convexity).

All of these empirical patterns are usually (mis)interpreted as a deviation from *the one and only* "correct" model and perceived as detrimental cognitive biases (Gigerenzer 2018; Lopes 1991). Thus the mode of explanation in behavioural economics is to utilise psychological biases and cognitive processes in the brain which cause certain empirical regularities, *e.g.* inverse-S curve labelled as probability weighting. From a normative point of view there remains the desideratum to provide a mechanistic explanation for each of these patterns. Our contribution is that we provide a mechanistic explanation for probability weighting and show that it may even be adaptive in the likely case of cautiously using a model with more uncertainty relative to another model.

Sunstein (2020) is only the latest paradigmatic example of this (mis)interpretation and points to the potential misguidance of this stream of literature, let alone design policy measures derived from this view.

****MK: End on a positive note MK****

References

Abdellaoui, M., Baillon, A., Placido, L., and Wakker, P. P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation. 101:695–723.

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, 27:173–196.

Gigerenzer, G. (2018). The bias bias in behavioral economics. *Review of Behavioral Economics*, 5:303–336.

Lattimore, P. K., Baker, J. R., and Witte, A. D. (1992). Influence of probability on risky choice: A parametric examination. *Journal of Economic Behavior and Organization*, 17:377–400.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168.

Lopes, L. L. (1991). The rhetoric of irrationality. 1:65–82.

Peters, O. (2019). The ergodicity problem in economics. *Nature Physics*, 15:1216–1221.

Prelec, D. (1998). The probability weighting function. 66:497–527.

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. 32:101–130.

Sunstein, C. R. (2020). The cognitive bias that makes us panic about coronavirus.

Tversky, A. and Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, 102(2):269–283.

Tversky, A. and Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. 5:297–323.

Tversky, A. and Wakker, P. (1995). Risk attitudes and decision weights. *Econometrica*, 63(6):1255–1280.

Wakker, P. (2010). *Prospect Theory*. Cambridge University Press.

Wu, G. and Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, 42:1676–1690.