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

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

April 4, 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).

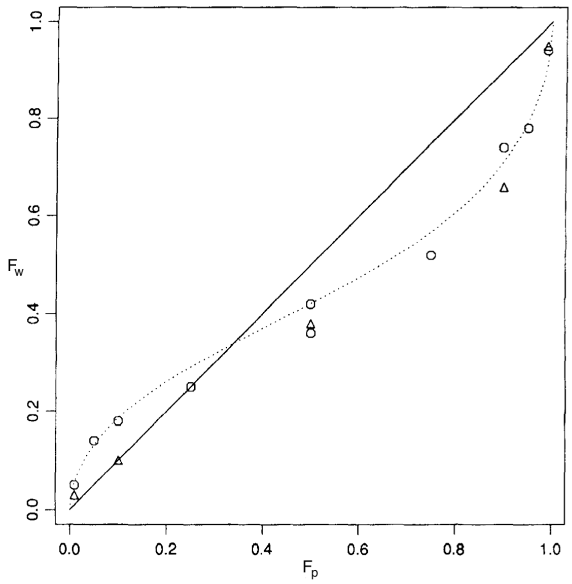

Figure 1: **Empirical phenomenon of probability weighting.** Cumulative decision weights F_w (used by decision makers) versus cumulative probabilities F_p (used by disinterested observers), as reported by (Tversky and Kahneman 1992, p. 310, Fig. 1). The figure is to be read as follows: pick a point along 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

We begin by making 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.

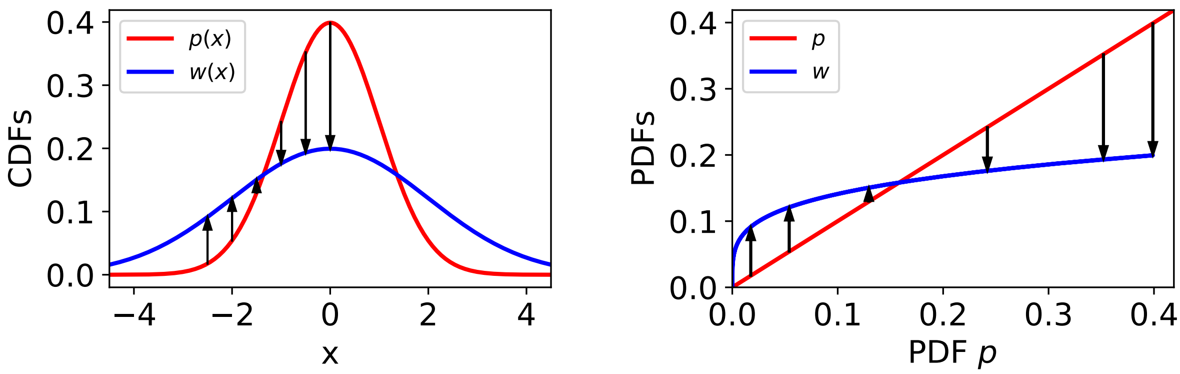

Figure 2: **Mapping PDFs.** Left: probability PDF (blue), estimated by a DO; and decision-weight PDF (red), estimated by a DM. The DO models x with a best estimate for the scale (standard deviation) and assumes the true frequency distribution is the blue line. The DM models x with a greater scale (here 2 times greater, $\alpha = 2$), and assumes the true frequency distribution is the red 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.

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

$$w(x) = \frac{1}{\sqrt{2\pi\alpha^2\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2(\alpha^2\sigma^2)}\right] \tag{4}$$

and

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

solve (Eq. 5) for $(x-\mu)^2$, substitute that in in (Eq. 4), and obtain the following expression for decision weights directly as a function of probabilities

$$w(p) = p^{\frac{1}{\alpha^2}} \frac{(2\pi\sigma_1^2)^{\frac{1-\alpha^2}{2\alpha^2}}}{\alpha} , \tag{6}$$

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.

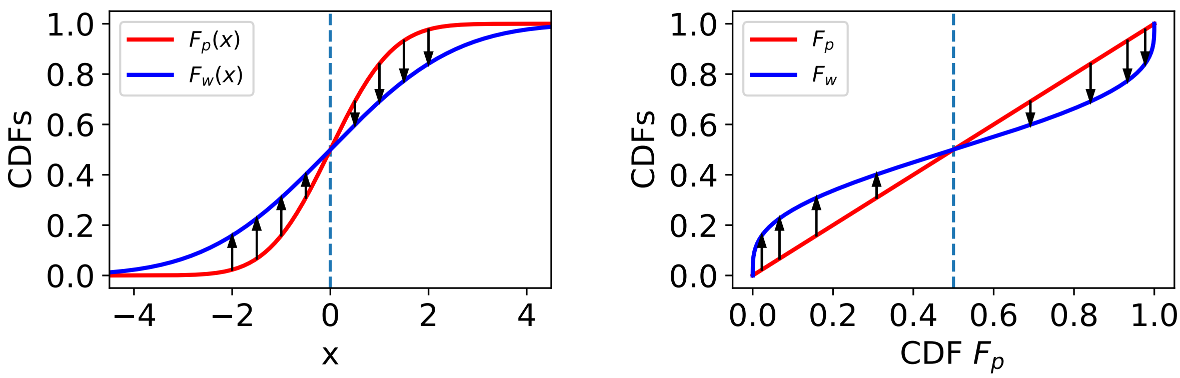

Figure 3: **Mapping CDFs.** Left: The DO assumes the observable X follows Gaussian distribution $X \sim \mathcal{N}(0,1)$, which results in the red CDF of the standard normal, $F_p(x) = \Phi_{0,1}(x)$. The DM is more cautious, in his model the same observable X follows a wider Gaussian distribution, $X \sim \mathcal{N}(0,3)$ 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.

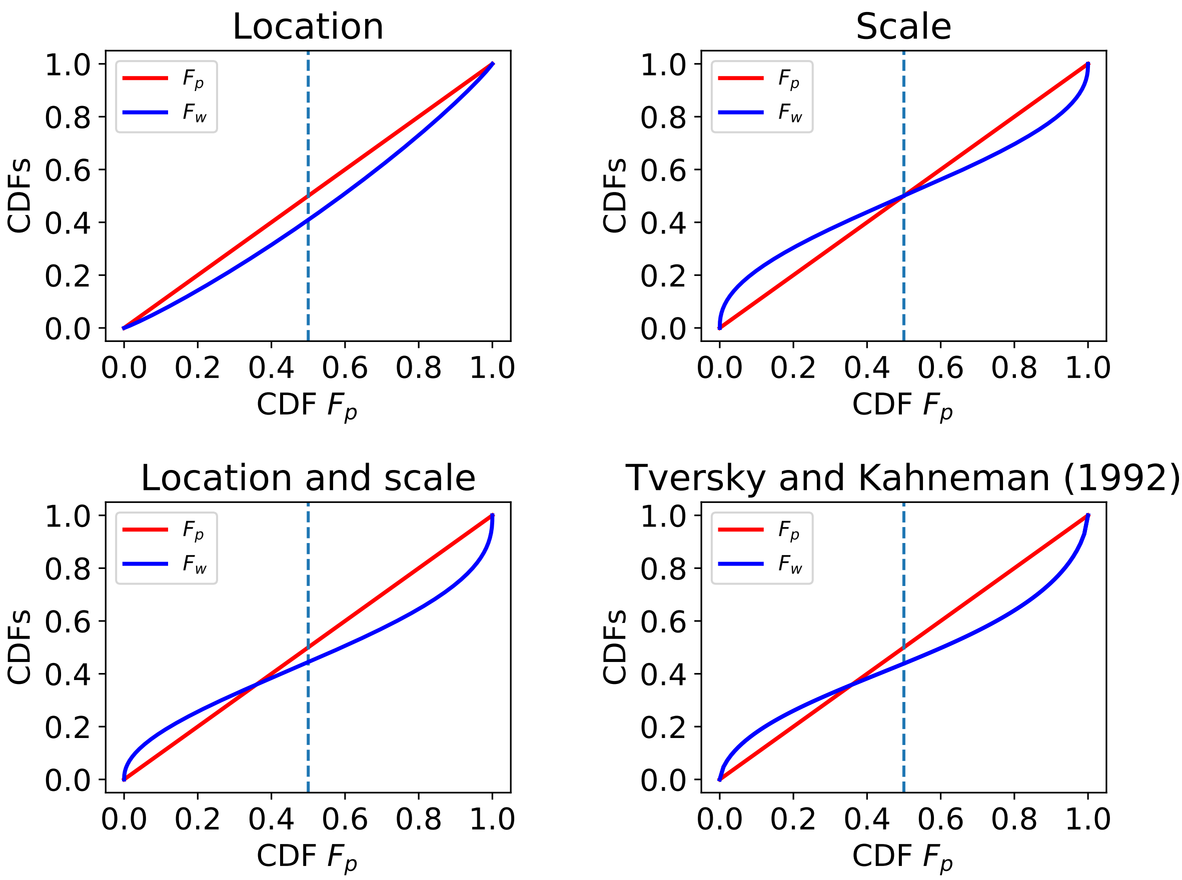

Figure 4: **Decision weight CDFs used by a DM vs. probability CDFs used by a DO, Gaussian distribution**.
Top left: Difference in scale. DO assumes location 0, scale 1; DM assumes location 0, scale 1.64 (broader than DO).
Top right: Difference in location. DO assumes location 0, scale 1; DM assumes location 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. 7) 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. The result is qualitatively similar to the bottom right panel of Fig. 3, corresponding to (Eq. 7).

It is telling that assuming only a difference in scale and location (for the simplest Gaussian case) is already sufficient to reproduce the observations that are characterised an irrational bias and are labelled as "probability weighting."

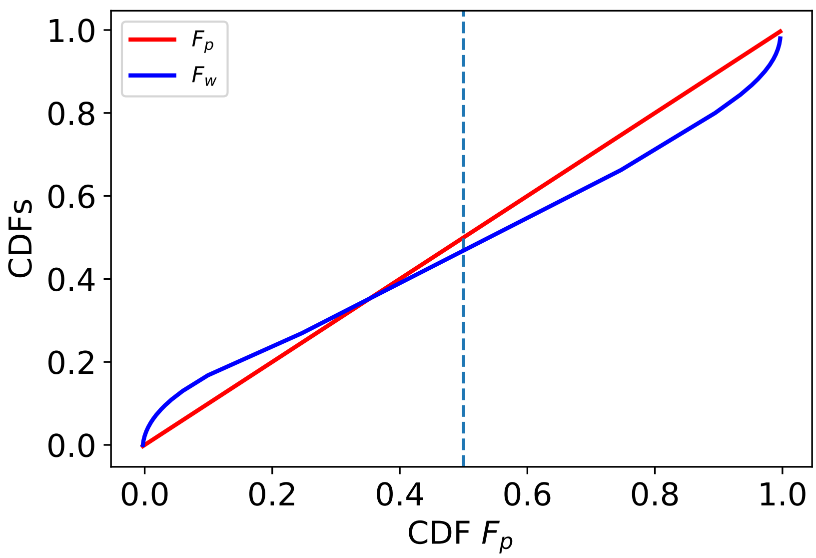

Figure 5: Probability weighting for Student-t distributions, where the DM uses a different shape parameter (1) and a different location parameter (0) from those of the DO (2 and 0.2, respectively).

3 Reasons for different models

We have put forward the possibility that probability weighting is not a detrimental cognitive bias but an obscure way to express that DMs tend to assume a greater range of plausible outcomes than DOs. For this to be a likely explanation of the phenomenon, 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 extra uncertainty. For example, even if the DO may know the true probabilities of some gamble in an experiment, and builds a formal model accordingly; the DM may in addition have doubts about the DO's sincerity, or about his understanding of the rules of the game. We will return to this in Sec. 3.2.

3.1 Lack of conceptual clarity about whose probabilities are analysed?

The conundrum surrounding the observed phenomenon of the characteristic inverse-S-shaped probability weighting vanishes if probability is understood as individual ignorance, *i.e.* somebody's uncertainty about an observable.

Frequency-in-an-ensemble interpretation of probability It's not easy to unpack a simple probability statement like "the probability of rain here tomorrow is 70%." Tomorrow only happens once, so rain happens in 70% of what? The technical answer to this question is usually: rain happens in 70% of the members of an ensemble of computer simulations of what may happen tomorrow run by a weather service. So one interpretation of "probability" is "relative frequency in a hypothetical ensemble of possible futures."

How exactly such a statement is linked to physical reality is not completely clear. Sometimes ensembles are real, for instance, when we say the probability of having a car accident is 1% per 10000 km driven – that's a summary of statistics collected over a large ensemble of cars. In this case, it's a real ensemble that existed in the past, not an imagined one in the future.

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.

Degree-of-belief interpretation of probability No matter how "probability" relates to a frequentist physical statement (whether with respect to an ensemble of simultaneous possibilities or to a time series), 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."

Many long books have been written about the meaning of probability; for our purpose it suffices to say that there's no guarantee that a probabilistic statement will be interpreted by the listener (DM) as it was intended by whoever (DO) made the statement.

Estimation errors for probabilities

Let's imagine the DO and DM have agreed explicitly on an interpretation of the word "probability." Say they agree that they mean the relative frequency in a long time series. Real time series are, of course, of finite length. In order to estimate the relative frequency of some event in a time series, we basically count – out of T time intervals, the event i occurred in n_i of them, so our best estimate for the probability is $\frac{n_i}{T}$. In the simplest case (and we rarely consider anything more complicated), we model the arrival of events as a Poisson process, where the standard error in the count of an event famously goes as $\sqrt{n_i}$. The standard error in the probability of an event is therefore $\frac{\sqrt{n_i}}{T}$, and the relative error is $\frac{1}{\sqrt{n_i}}$. Low probabilities therefore come with larger relative errors, see Tab. 1. We note that behaviourally, it will make little difference whether a DM assigns a 0.49 probability to an event or a 0.51 probability. It will make a large difference, however, whether a DM assigns a 0.002 probability or a 0.0002 probability.

The most important message from this example is that errors in probability estimates behave differently for low probabilities than for high probabilities: absolute errors are smaller for low probabilities, but relative errors are larger for lower probabilities.

Asymptotic (true) probability	Most likely count	Absolute estimation error	Relative estimation error
.5	5000	.01	2%
.1	1000	.003	3%
.01	100	.001	10%
.001	10	.0003	30%
.0001	1	.0001	100%

Table 1: This table assumes $T = 10000$ observed time intervals. To be read as follows (first line): for an event of true probability 0.5, the most likely count in 10000 trials is 5000. Assuming Poisson statistics, this comes with an estimation error of $\sqrt{5000}/5000 = 0.01$, which is 2% of the true probability.

3.2 Disinterested observers and decision makes have different perspectives

****MK: find a better section title MK****

The observations that led to the concept of probability weighting in prospect theory can be expressed as follows: DOs assign systematically lower weights to low-probability events than DMs. Which of the two is wrong is unclear so long as it is unclear who means what by the word "probability." Because the two types of modellers (DO and DM) pursue different goals, it may be the case that neither is wrong about the probabilities, just wrong about the goals of the other modeller.

Being a good neutral scientist, a DO has no particular interest in the success or failure of a DM. Being a good DM, the DM has every interest in his own success or failure. 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 *i.e.* their wealths) necessary for evolution ****MK: noise generates mutations in the genotype, the noise in the world is a different animal IMO MK**** and multiplicative growth (self-reproduction) is how successful phenotypes spread their traits in a population ****MK: Covid-19 reference tempting MK****. This process operates on amoeba, as it does on forms of institutions, and on investment strategies.Peters and Adamou (2018)

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, and underestimating it would have catastrophic consequences, whereas overestimating it may lead the DM to miss out on some opportunities.

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.

In Fig. 4 we show maps that result from the procedure illustrated in Fig. 3. We show separately maps that arise when the DM models the scale of the distribution differently (see Fig. **??**), and when the DM models the location differently (see Fig. **??**). We then put the two together (see Fig. **??**) and compare the result to the functional shape of F_w^* which Tversky and Kahneman (1992) chose to fit to their observations (see Fig. **??**).

Without any derivation from a physical mechanism which would motivate the specific functional form, Tversky and Kahneman (1992) chose to fit the following function to resemble their data[2]

$$F_w^*\left(F_p^*\right) = \left(F_p^*\right)^\gamma \frac{1}{\left[\left(F_p^*\right)^\gamma + \left(1 - F_p^*\right)^\gamma\right]^{1/\gamma}} \,. \tag{7}$$

The function has only one free parameter, γ. For $\gamma = 1$ both functions collapse $F_w^*\left(F_p^*\right) = F_p^*$. The function F_w^* has the following property: any curvature moves the intersection with the diagonal away from the mid-point $1/2$. For this asymmetry to be reproduced in

[2]Equation (7) is the consensus functional form in the community Barberis (2013). Whereas we provide a mechanistic explanation, psychological explanations prevail in the behavioural economics and finance literature, see Abdellaoui et al. (2011); De Giorgi and Hens (2006); Gonzalez and Wu (1999); Prelec (1998); Stott (2006); Wakker (2010); Wu and Gonzalez (1996) and references especially in the latter. It remains doubtful how a psychological explanation shall compensate for a missing and/or not yet understood mechanism, generating a particular phenomenon like probability weighting.

the Gaussian case, it is necessary to introduce a difference between the both the locations used by DO and DM in addition to the difference between scales (see Fig. **??**). If we allow the possibility that the DM uses different estimates for scale and location, we can reproduce the observations in Tversky and Kahneman (1992) accurately, see Fig. 4.

4 Fitting the model to experimental results

To test whether our model is in line with existing evidence on probability weighting, we fit the Gaussian and Student's-t distributions to experimental data. Specifically, we fit the location (μ) and scale (σ) parameters in the Gaussian case and the shape parameters (ν and δ) in the Student's-t case to experimental results from Tversky and Kahneman (1992) (depicted in circles in Fig. 1) and from Tversky and Fox (1995). In addition, we fit the function

$$w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}}\,, \tag{8}$$

suggested by Lattimore et al. (1992) to parametrically describe probability weighting, and was commonly used since (see, *e.g.* Tversky and Wakker (1995)).

Figure 6 presents these results. We obtain very good fit to data for both Gaussian and student's-t distributions, as well as for (Eq. 8), in the two experiments. It is practically impossible to distinguish between the three fitted functions within standard errors. We conclude that our model fit the data well, and unlike (Eq. 8), the fitted functions are directly derived from a mechanistic model, and are not simply phenomenological.

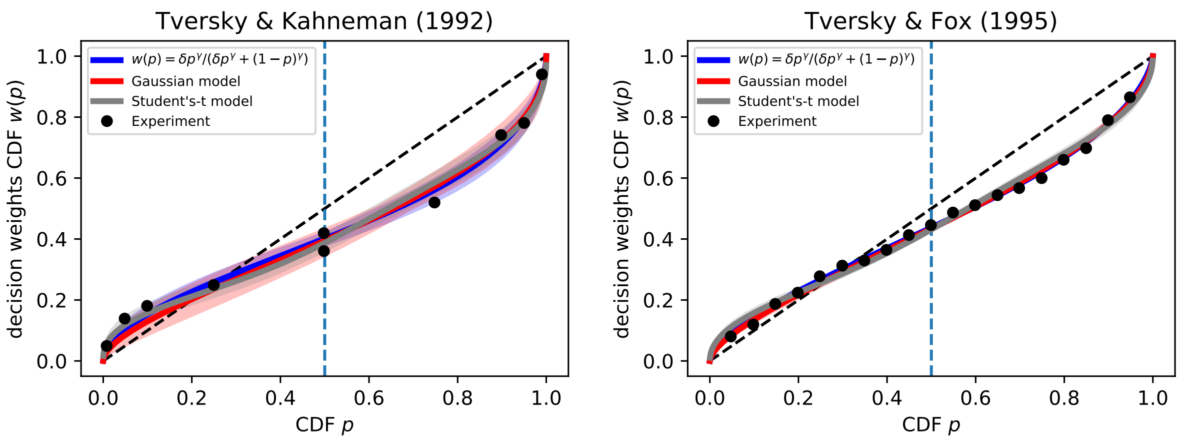

Figure 6: Model fitting to experimental data from Tversky and Kahneman (1992) (left) and Tversky and Fox (1995) (right). Left) Gaussian model: $\mu = 0.4$ ($SE = 0.13$), $\sigma = 1.5$ (± 0.2); Student's-t model: $\nu = 0.39$ (± 0.06), $\delta = 0.29$ (± 0.08); Lattimore et al. (1992): $\delta = 0.67$ (± 0.07), $\gamma = 0.59$ (± 0.06). Right) Gaussian model: $\mu = 0.23$ (± 0.03), $\sigma = 1.4$ (± 0.06); Student's-t model: $\nu = 0.47$ (± 0.05), $\delta = 0.17$ (± 0.03); Lattimore et al. (1992): $\delta = 0.77$ (± 0.02), $\gamma = 0.69$ (± 0.02). Shaded areas (noticeable only in the case of Tversky and Kahneman (1992)) indicate one standard error in the fitted parameter values. The fit was done by implementing the method of least squares with the Nelder-Mead algorithm Nelder and Mead (1965), and the standard errors were obtained by bootstrapping.

5 Conclusion/Summary

****MK: Discuss the general action of change of measure in relation to Girsanov and risk-neutral prob measure $\mathbb{Q}$ in Math Finance? MK****

****MK: It has almost become custom to end with a Kacelnik quote 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.

De Giorgi, E. and Hens, T. (2006). Making prospect theory fit for finance. 20:339–360.

Gonzalez, R. and Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38:129–166.

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.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4):308–313.

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

Peters, O. and Adamou, A. (2018). An evolutionary advantage of cooperation. *arXiv:1506.03414*.

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

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

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.