Probability weighting as model calibration

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

March 6, 2020

Abstract

Behavioral economics collects observations of human economic behavior and provides labels for those observations. Probability weighting is one such label. It expresses a mismatch, in decision problems, between probabilities used in a formal model of a decision problem (*i.e.* model parameters) and probabilities inferred from real people's behavior 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 normalization) lower than probabilities for common events. The observed behavior 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 model well-specified reasons for such caution and find the resulting probability weighting, as a benchmark for reasonable behavior. We find close agreement with empirical studies.

Keywords: decision theory, prospect theory, probability weighting, ergodicity economics

[*]London Mathematical Laboratory and Santa Fe Institute, o.peters@lml.org.uk
[†]London Mathematical Laboratory, a.adamou@lml.org.uk
[‡]London Mathematical Laboratory, m.kirstein@lml.org.uk
[§]London Mathematical Laboratory, y.berman@lml.org.uk

1 Nomenclature and key observation

Probability weighting originates in prospect theory (Barberis, 2013). 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]

This key observation is often summarized visually with a comparison between

- cumulative probability density functions (CDFs), denoted

$$F_p(x) = \int_{-\infty}^{x} p(s)ds \tag{1.1}$$

  and

- decision weight CDFs, denoted

$$F_w(x) = \int_{-\infty}^{x} p(s)ds. \tag{1.2}$$

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

As a final piece of nomenclature, we will use the terms *location* and *scale* when discussing probability distributions. Consider a standard normal distribution $\mathcal{N}(0, 1)$ – here, the parameters indicate

[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 normalizing just means dividing by a constant (the sum or integral of the non-normalised decision weights).

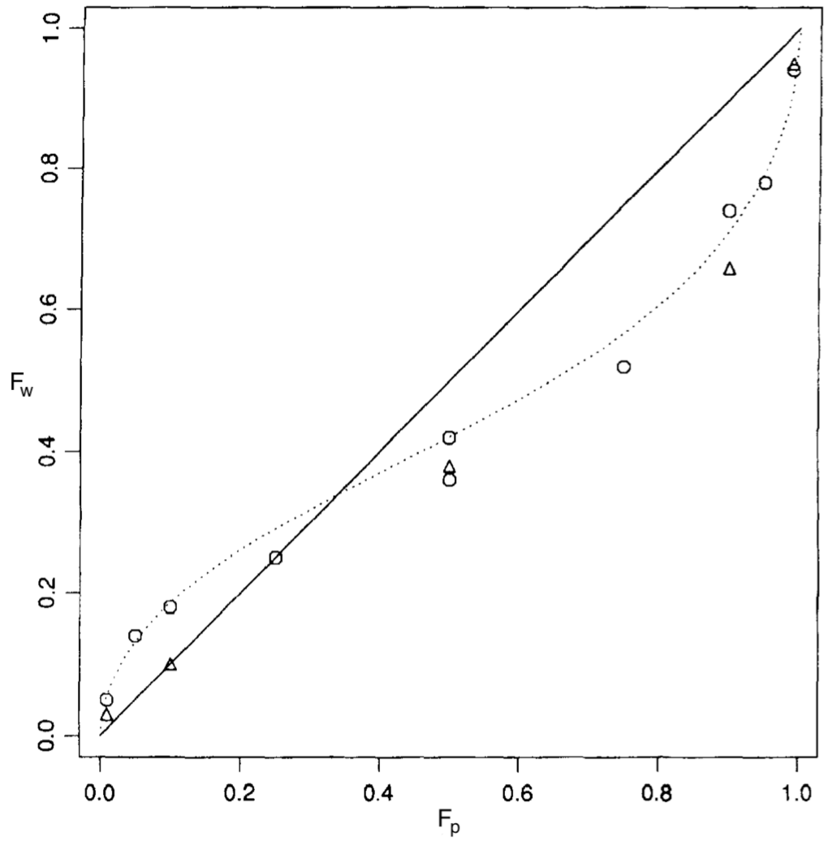

Figure 1: Cumulative decision weights F_w (used by decision makers) versus cumulative probabilities F_p (used by disinterested observers), as reported by Tversky and Kahneman (1992). The figure is to be read as follows: pick a point along the x axis (the probability used by a DM) and look up the corresponding value on the y axis (the decision weight). For small (cumulative) probabilities (left) the decision weight exceeds the probability, and for large probabilities it's the other way around. It's the inverse-S shape of the curve that indicates this qualitative relationship.

location 0 and squared scale 1 (for a Gaussian that's the mean and variance). In general, the identically-shaped distribution of location μ and scale σ is obtained from the location-0, scale-1 distribution by transforming the argument x according to

$$y = \frac{x - \overbrace{\mu}^{\text{location}}}{\underbrace{\sigma}_{\text{scale}}} \tag{1.3}$$

The PDF $p(y)$ has location μ and scale σ.

We will show below that these observations are predicted by considerations of uncertainty about probabilities, or more generally by uncertainty about model parameters.

2 Consistent probability weighting as a difference between models

Behavioral economics interprets Fig. 1 as evidence for a perception bias in the DM. We will keep a neutral stance. We don't consider the DO to know "the truth" – he has a model of the world. Nor do we consider the DM to know "the truth" – he has another model of the world. Figure 1 shows that the two models differ. Below we will go through a few generic reasons why the models may differ, and we will find that the inverse-S curve is a robust prediction in ergodicity economics – when we emphasize that DMs have to operate along time lines.

2.1 Generic case: the Gaussian example

Our key point is that the robust qualitative observation of the inverse-S shape in Fig. 1 is reproduced by assuming that the DM uses a larger scale in his model of the world than the DO. This can have numerous reasons, precaution being perhaps the most 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_1^2. 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 $\sigma_1^2 + \sigma_2^2$. The DM simply assumes a broader range of plausible values, Fig. 2.

Generically, the effect of the DM using a greater scale in his model is higher decision weights for low-probability events, and (because of normalization), lower decision weights for high-probability events.

In the Gaussian case we can write the distributions explicitly

$$w = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left[\frac{-(x-\mu)^2}{2(\sigma_1^2 + \sigma_2^2)}\right] \tag{2.1}$$

and

$$p = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma_1^2}\right]. \tag{2.2}$$

Furthermore, we can solve Eq. (2.2) for $(x-\mu)^2$ and substitute that in in Eq. (2.1). Thereby, we obtain the following expression for decision weights directly as a function of probabilities

$$w(p) = p^{\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}} \frac{\left(\sqrt{2\pi\sigma_1^2}\right)^{\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}}}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}, \tag{2.3}$$

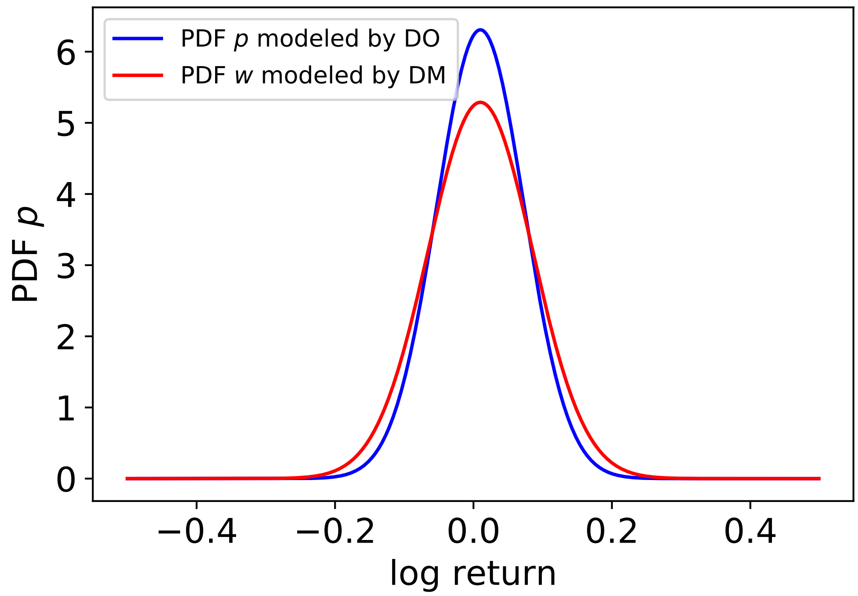

Figure 2: Probability density function (blue), estimated by a DO; and decision-weight density function (red), estimated by a DM. The DO models returns with a best estimate for the variance and assumes the true frequency distribution is the blue line. The DM wants to be on the safe side, models returns with a greater variance, and assumes the true frequency distribution is the red line. The DM appears to the DO as someone who over-estimates probabilities of low-probability events and underestimates probabilities of high-probability events.

which we plot in Fig. 3.

In Fig. 4 we plot the corresponding CDFs, F_w against F_p. This mechanism does not, of course, rely on the Gaussian assumption: the inverse-S shape is found whenever the DM assumes a greater scale for a unimodal distribution.

We have put forward the possibility that probability weighting is not cognitive bias but a complicated way to say 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, Sec. 2.2.

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 just 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, in an experiment, the DO may know the

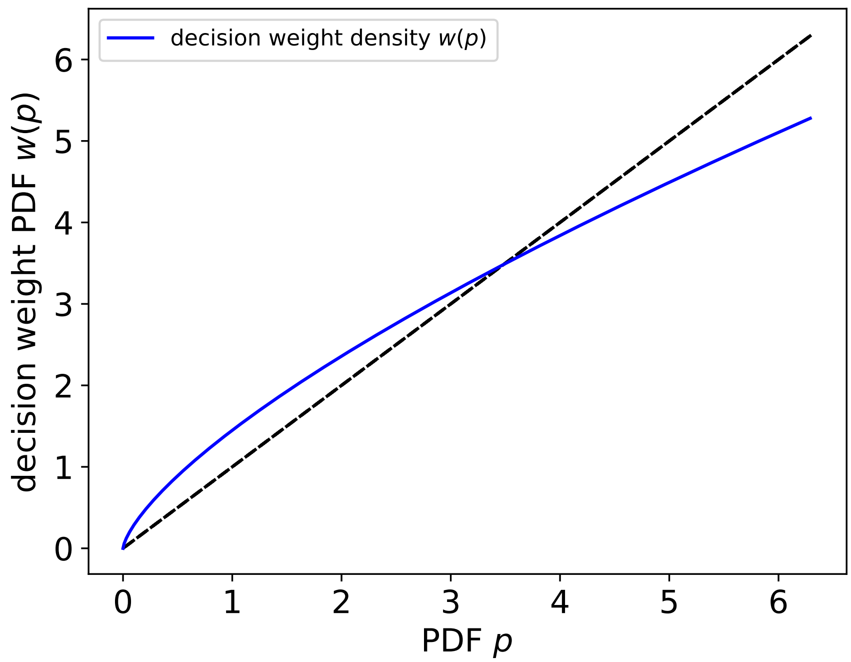

Figure 3: Decision weight density (used by a DM) vs. probability density (used by a DO) for the Gaussian model (blue), compared to the diagonal (black) where DM and DO use the same parameters. For low probabilities, the decision weights are higher than the probabilities; for high probabilities they are lower.

true probabilities of some gamble and build a forma 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. 2.3.

2.2 Probabilities are tricky

Lack of conceptual clarity

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, 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."

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 10,000km 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.

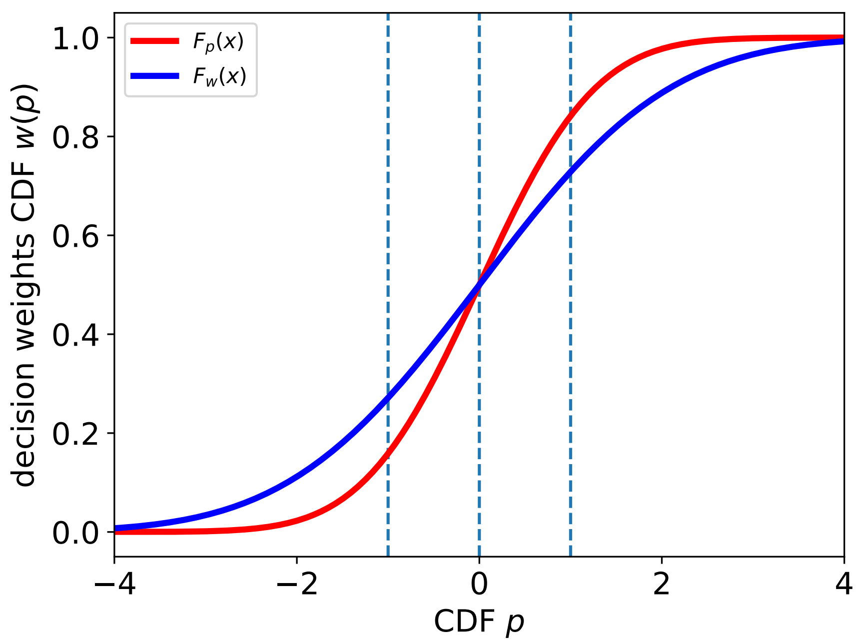

Figure 4: CDF as assumed by the DO (red) and by the DM (blue), Gaussian distributions, where the DO assumes scale 1 and the DM cautiously assumes scale 2. Following the dashed vertical lines (left to right), we see that for small values of the CDF the DM's is larger than the DO's; the curves coincide at 0.5 because no difference in location is assumed; and for large values of the CDF the DM's is smaller than the DO's

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.

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 as it was intended by whoever 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 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 probability or a 0.0002 probability. The most important message from this example computation is that errors in probability estimates behave differently for small probabilities than for large probabilities: absolute errors are smaller for small probabilities, and relative errors are larger for small probabilities.

Table 1: This table assumes T =10,000 observed time intervals. To be read as follows (first line): for an event of true probability 0.5, the most likely count in 10,000 trials is 5,000. Assuming Poissonian statistics, this comes with an estimation error of $\sqrt{5,000}/10,000 = 0.01$, which is 2% of the true probability.

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

2.3 Disinterested observers and decision makes have different perspectives

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 the success or failure of the DM. 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, 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 small-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.

In Fig. 5 we show maps that result from the procedure illustrated in Fig. 4. We show separately maps that arise when the DM models the scale of the distribution differently, and when the DM models the location differently. We then put the two together and compare the result to the functional shape Tversky and Kahneman (1992) chose to fit to their observations.

Without a mathematical derivation or physical motivation of the functional form, Tversky and Kahneman (1992) chose to fit the following function to resemble their data

$$F_{w_{TK}}(F_{p_{TK}}) = F_{p_{TK}}^{\alpha} \frac{1}{(F_{p_{TK}}^{\alpha} + (1 - F_{p_{TK}}))^{1/\alpha}}. \tag{2.4}$$

This function only has one free parameter, α, and has the following property: any curvature (*i.e.* any deviation from $F_{w_{TK}}(F_{p_{TK}}) = F_{p_{TK}}$, which is the shape it takes for $\alpha = 1$) moves the intersection with the diagonal away from the mid-point $1/2$. For this feature to be reproduced in the Gaussian case, it is necessary to introduce a difference between the locations used by DO and DM in addition to the difference between scales. 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. 5.

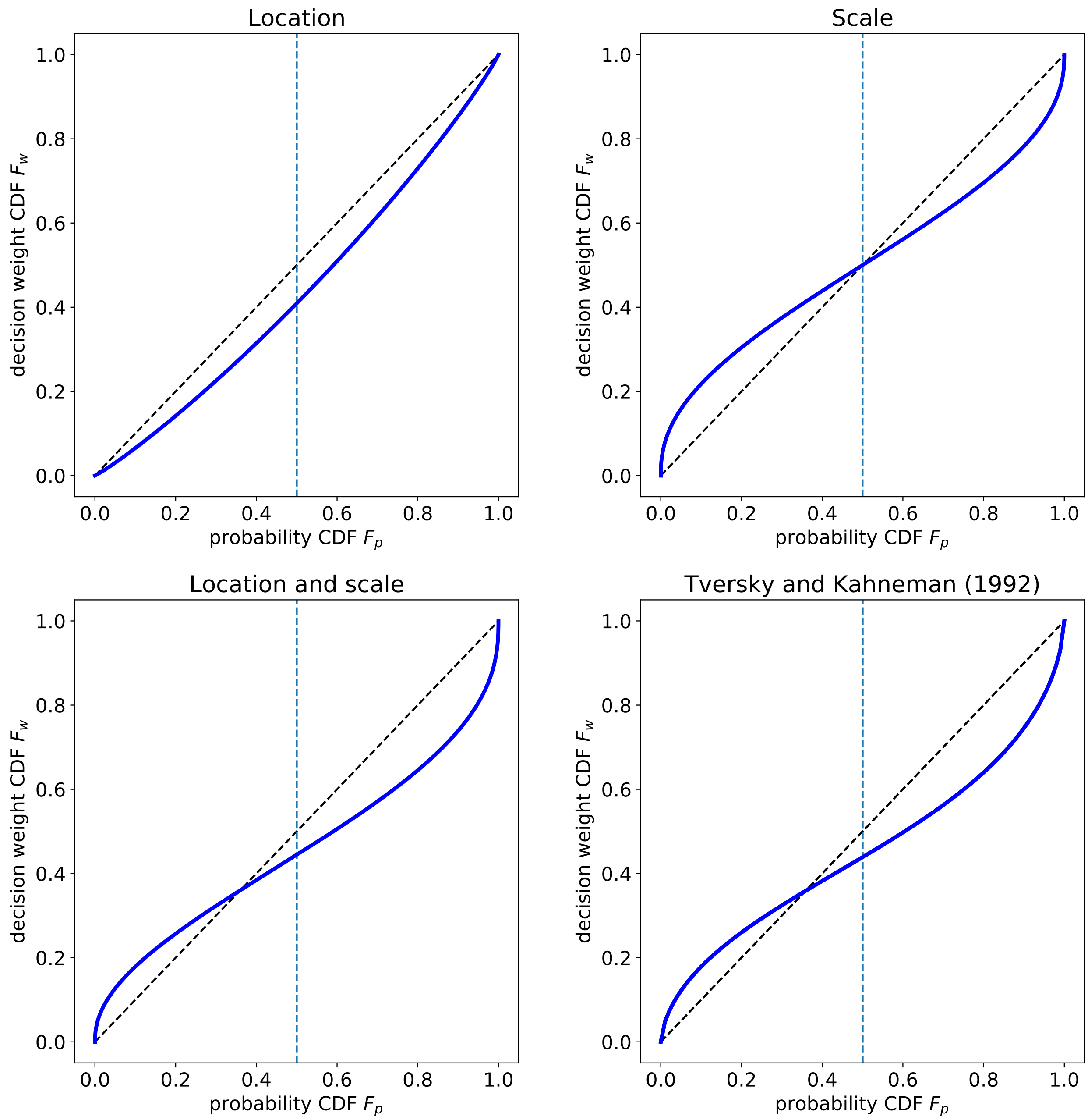

Figure 5: Decision weight CDFs used by a DM vs. probability CDFs used by a DO.
Top left) Gaussian distribution, difference in scale. DO assumes location 0, scale 1; DM assumes location 0, scale 2.7 (broader than DO).
Top right) Gaussian distribution, difference in location. DO assumes location 0, scale 1; DM assumes location 0.18 (bigger than DO), scale 1. Bottom left) Gaussian distribution, differences in scale and location. DO assumes location 0, scale 1; DM assumes location 0.18 (bigger than DO), scale 2.7 (broader than DO).
Bottom right) Fit to observations reported by Tversky and Kahneman (1992). This is Eq. (2.4) with $\alpha = 0.65$. The observations by Tversky and Kahneman (1992) are consistent with a DM assuming a scale and location in real-world decisions that differ from those assumed by the DO.

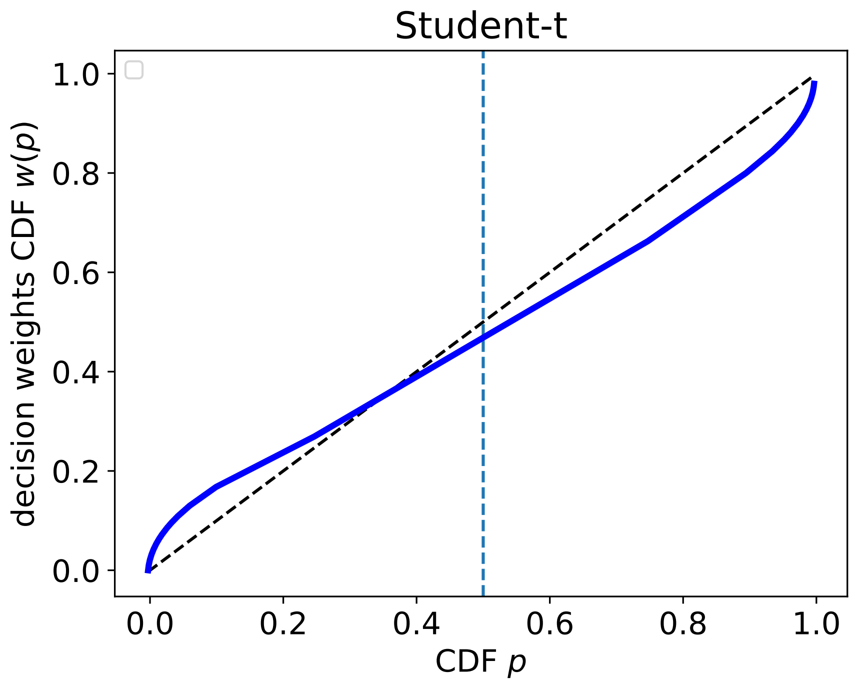

Figure 6: Probability weighting for Student-t distributions, where the DM uses a different shape parameter (1) and a different location parameter (0) from those of the DO (2 and 0.2, respectively).

3 Other probability 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 assume a type of distribution that differs from the DO's. An infinity of combinations of assumed distributions can be explored. To illustrate the generality of the procedure, we carry it out for a (power-law tailed) Student-t distribution, Fig. 6, where DO and DM use different shape parameters and different locations. The result is qualitatively similar to Fig. 5 D).

It is telling that assuming only a difference in scale and location, for the simplest case (Gaussian) reproduces the observations that have come to be known as "probability weighting."

References

Barberis, N. C. (2013): "Thirty Years of Prospect Theory in Economics: A Review and Assessment," Journal of Economic Perspectives, 27, 173–96.

Peters, O. (2019): "The Ergodicity Problem in Economics," Nature Physics, 15, 1216–1221.

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