Information Theory Meets Expected Utility: The Entropic Roots of Probability Weighting Functions

Mikaela Akrenius (makreniu@indiana.edu)
Cognitive Science Program, Indiana University Bloomington
1900 E. 10th Street, Bloomington, IN 47406 USA

Abstract

This paper proposes that the shape and parameter fits of existing probability weighting functions can be explained with sensitivity to uncertainty (as measured by information entropy) and the utility carried by reductions in uncertainty. Building on applications of information theoretic principles to models of perceptual and inferential processes, we suggest that probabilities are evaluated relative to a plausible expectation (the uniform distribution) and that the perceived distance between a probability and uniformity is influenced by the shape (relative entropy) of the distribution that the probability is embedded in. These intuitions are formalized in a novel probability weighting function, *VWD*(p), which is simpler and has less parameters than existing probability weighting functions. The proposed probability weighting function captures characteristic features of existing probability weighting functions, introduces novel predictions, and provides a parsimonious account of findings in probability and frequency estimation related tasks.

Keywords: decision making under risk and uncertainty; probability weighting; information entropy; predictive coding

Introduction

We are faced with hundreds of decisions every day. These choices range from the seemingly simple and trivial, such as whether to cross a busy street or not, to the more complex and consequential, such as which job offer to accept or which public policy to vote on. Regardless of their domain, they all share a common characteristic: an assessment of the likelihood and qualitative nature of possible future outcomes.

A classical framework for studying how people make choices that involve uncertain outcomes is the *expected utility maximization* paradigm. In the paradigm, participants are asked to choose between monetary gambles that each consist of a set of mutually exclusive outcomes and their respective probabilities. For example, a participant might be asked to choose between 'Gamble A: $100 with certainty' or 'Gamble B: $200 with a 50% chance, $0 otherwise.' When presented sequentially with varied outcome magnitudes and probability distributions, responses to these choice problems are presumed to reveal inner preference orderings between outcomes and attitudes to risk and uncertainty. Following the *expected value maximization principle* introduced by Bernoulli (1738/1954) and axiomatized by von Neumann and Morgenstern (1944/1947) as *expected utility theory* (EUT), participants are assumed to choose the gamble that has the highest expected utility (EU):

$$EU(X) = \sum_{i=1}^{n} p_i * u(x_i) \tag{1}$$

where X denotes a gamble, x_i outcomes of the gamble, $u(x_i)$ their utilities, p_i their respective probabilities, and n the number of outcomes in the gamble.

Later studies (most notably Allais, 1953, and Ellsberg, 1961; see also Camerer, 1989, for a review) have shown that, with an appropriate framing of the choice problem, people can be induced to make choices that violate the axioms of EUT. To account for this, a plethora of variations of EUT (termed *non-expected utility theories*) have been proposed (see e.g. Machina, 2008, for a review) that aim to account for the anomalies through adjusting the shape of the utility function and/or incorporating a *probability weighting function*, yielding

$$EU(X) = \sum_{i=1}^{n} w(p_i) * u(x_i) \tag{2}$$

where $w(p_i)$ denotes the weighted probability p_i.

Most notably, *prospect theory* (Kahneman & Tversky, 1979) presumes a value (utility) function that is concave for gains, convex for losses, and steeper for losses than for gains, and a probability weighting function that overweights small probabilities and underweights larger probabilities. Later major applications of EUT (*rank-dependent utility theory*, Quiggin, 1982, and *cumulative prospect theory*, Tversky & Kahneman, 1992) suppose that probabilities are weighted *cumulatively*, i.e. that the outcomes of each gamble are multiplied with the weighted probability of receiving that outcome *or more* (gains) or that outcome *or less* (losses).

Even though these theories are widely applied and can account for many findings that EUT or non-cumulative prospect theory cannot account for (Camerer, 2000; Machina, 2008), recent research has introduced new sets of anomalies that are unexplained by rank-dependent utility theories (e.g. Birnbaum, 2006, 2008; Birnbaum et al., 2008). Some researchers in the field have proposed that novel expected utility based frameworks should be introduced (e.g. Luce, 2008), and that the psychological underpinnings of these models should be more firmly established (e.g. Kahneman, 2003).

This paper suggests that the notion of information entropy and its postulated utility in perceptual and inferential processes could provide a theoretical framework that can explain properties of non-expected utility theories, introduce new empirical predictions, and give rise to new modeling paradigms. This approach is exemplified in a novel entropy-based probability weighting function that does the above.

Probability Weighting as Sensitivity to Deviations from Maximum Entropy

What Is Probability Weighting?

Psychologically, a *probability weighting function* describes distortions in the experience of a subjective degree of belief (i.e. subjective probability). In the context of decision making under risk, subjective probabilities are presumed to equal the numerical probabilities presented to the participant, and choosing as if the outcome attached to a probability were more/less likely than it actually is is interpreted as *over-/underweighting* the probability. Because probabilities p are restricted to the range [0, 1], a probability weighting function $w(p)$ is restricted to the same range, defining $w(0) = 0$ and $w(1) = 1$. Based on accumulated experimental evidence (see e.g. Kahneman & Tversky, 1979; Gonzalez & Wu, 1999), small probabilities are generally overweighted ($w(p) > p$) and larger probabilities underweighted ($w(p) < p$), yielding an inverse-S-shaped, continuous, and monotonically increasing function.

Existing Probability Weighting Functions

Up to date, a plethora of probability weighting functions have been proposed (see e.g. al-Nowaihi & Dhami, 2010; Cavagnaro et al., 2013; and Stott, 2006), of which the *Tversky-Kahneman* weighting function (Tversky & Kahneman, 1992), the empirically derived *linear-in-log-odds* weighting function (Goldstein & Einhorn, 1987; Tversky & Fox, 1995; Gonzalez & Wu, 1999), and the axiomatically derived *Prelec* weighting function (Prelec, 1998) generally provide best fits with empirical data (al-Nowaihi & Dhami, 2010; Cavagnaro et al., 2013; Stott, 2006). All of these functions share the properties listed above. They can differ, however, in their *curvature* (magnitude of over- or underweighting), *elevation* (amount of general over- or underweighting), *fixed point* $w(p) = p$, below which p is overweighted and above which p is underweighted, and *inflection point* $w(p)'' = 0$, where the curvature of the function changes from concave to convex. These properties can be adjusted through parameters.

Figure 1 depicts examples of weighting functions acquired using a range of parameter value combinations. The 1-parameter-Prelec weighting function only controls the curvature α of the weighting function, whereas the linear-in-log-odds and 2-parameter-Prelec weighting functions also adjust the elevation β of the weighting function. The Tversky-Kahneman weighting function controls both with one parameter, γ. As illustrated by the figure, adding a parameter increases the flexibility of the weighting function.

Applications of Entropy in Modeling Perceptual and Inferential Processes

In other areas of experimental psychology, it has been shown that the perceptual organization of visual stimuli can be modeled as minimizing uncertainty or code length (see e.g. Attneave, 1954; Garner & Clement, 1963; or Feldman, 2016, for a more recent summary), and that information search can be modeled as maximizing the amount of uncertainty reduction that gaining a novel piece of information entails (Oaksford & Chater, 1994; Crupi et al., 2018). This suggests that reducing the amount of uncertainty carries utility both at the perceptual and at the cognitive level. Researchers have further suggested that perceptual processes may conform to the same principles that control reasoning under uncertainty (see e.g. Friston, 2010; Knill & Pouget, 2004; Summerfield & Tsetsos, 2015). Building on the conjecture that reducing uncertainty carries utility, the human cognitive system may be tuned to context-specific fluctuations in the uncertainty associated with frequencies or probabilities. In particular, if the brain were to encode deviations from predictions rather than all information present to conserve time and energy (as e.g. Bubic et al., 2010 suggest), probabilistic information could also be represented as a deviation from a plausible expectation rather than as an absolute magnitude. Furthermore, this representation could be influenced by the overall amount of uncertainty present in the choice context (cf. the influence of stimulus statistics on the noisiness of number representations, Prat-Carrabin & Woodford, 2020).

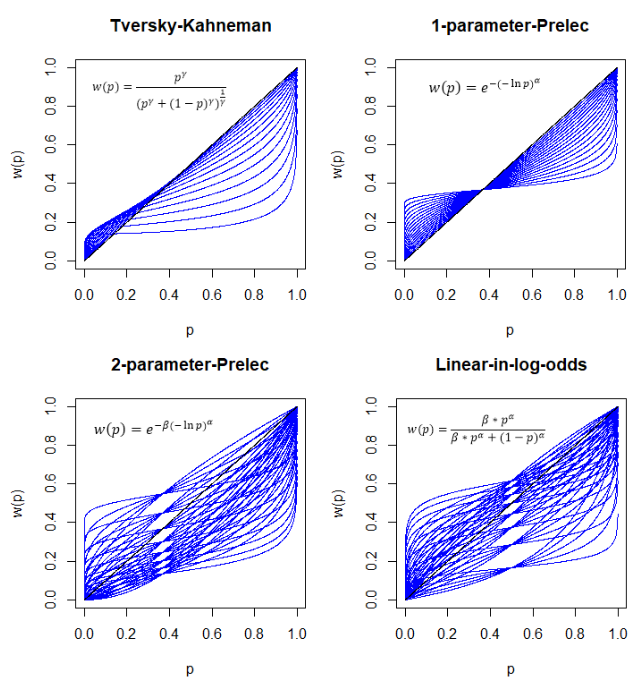

Figure 1: Shapes and functional forms of existing probability weighting functions.

An Information-Entropy Based Probability Weighting Function

In experiments of decision making under risk, a probability p is embedded in a gamble that consists of a distribution of n probabilities p_i that sum up to 1 and that are each associated with an outcome x_i. When no prior information related to the gamble is given, a plausible expectation for the shape of the probability distribution is uniformity, i.e. that $p_i = 1/n$ for all

p_i in the distribution. This is the distribution of maximum entropy. The notion of evaluating probabilistic information relative to a plausible expectation can be expressed as evaluating the distance $|p - 1/n|$. When p is associated with a positive outcome x and $p > 1/n$, $|p - 1/n|$ carries positive utility (because $p*x$ yields a larger expected value than $1/n*x$), whereas when $p < 1/n$, $|p - 1/n|$ carries negative utility (because $p*x$ yields a smaller expected value than $1/n*x$). In addition, the closer to $1/n$ (and further from 0 and 1) p and other probabilities in the distribution are, the more uncertainty is involved with predicting the outcome of the gamble. If an increase in uncertainty carries negative utility for positive expectations and positive utility for negative expectations (as findings by Bouchouicha, Martinsson, Medhin, and Vieider, 2017, would suggest), when $p > 1/n$, the "goodness" of $|p - 1/n|$ is tampered by the proximity of $1/n$, whereas when $p < 1/n$ the "badness" of $|p - 1/n|$ is lessened by the proximity of $1/n$. Hence, the closer p (and other probabilities in the distribution) are to $1/n$, the more the distance $|p - 1/n|$ is discounted. A similar treatment for negative outcomes $-x$ gives an equivalent result.

Taken together, these inferences can be formalized as

$$w(p) = \frac{1}{n} + \left(p - \frac{1}{n}\right) * \left(1 - \frac{H_{dist}}{H_{max}}\right) \quad (3)$$

where n denotes the number of probabilities in the distribution, H_{dist} refers to the Shannon entropy $[-\Sigma p*\log_2(p)]$ of the distribution, H_{max} is the entropy of the uniform distribution $[-\log_2(1/n)]$, H_{dist}/H_{max} is equivalent to the Kullback-Leibler divergence $D(H_{dist} \| H_{max})$ and defined in the range [0, 1], and $1 - H_{dist}/H_{max}$ is the redundancy of the distribution.

When combined with the requirement $w(0) = 0$, (3) can be applied to derive a probability weighting function

$$VWD(p) = \frac{p^{1-\frac{H_{dist}}{H_{max}}}}{\sum_{i=1}^{n} p_i^{1-\frac{H_{dist}}{H_{max}}}} \quad (4)$$

where *VWD* refers to *Valence-Weighted Distance* (the distance of p from $1/n$ weighted with x). Because H_{max} changes when n changes and H_{dist} changes when p (or any other probability in the distribution) changes, $VWD(p)$ is n-dimensional. The fixed point of $VWD(p)$ is at uniformity, i.e. at $p = 1/n$ for all p, and the curvature of the weighting function (i.e. amount of over- or underweighting of p) is determined by the relative entropy H_{dist}/H_{max} of the distribution. These two properties reflect the contextual influence of other probabilities on the evaluation of p and are illustrated in Figure 2. Because plotting $VWD(p)$ for n probabilities requires n dimensions, the functions depicted are samples from the entire $VWD(p)$.

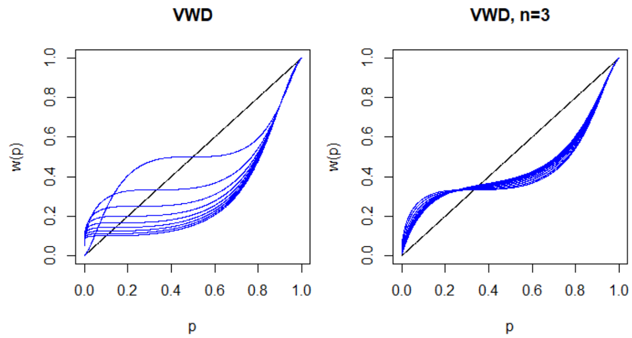

Figure 2: The influence of n (left panel) and distributional shape (right panel) on $VWD(p)$. The left panel shows the shape of $VWD(p)$ from $n = 2$ (highest curve) to $n = 8$ (lowest curve) when p is embedded in a distribution with $n - 1$ equal probabilities. The right panel illustrates the influence of the shape of the distribution $[p, (1-p)/j, (j-1)*(1-p)/j]$ on the curvature of $VWD(p)$ as j increases from 2 (most entropy) to 5 (least entropy). As the entropy of the distribution increases, the curvature of the weighting function also increases.

Comparing Properties of *VWD*(*p*) to Properties of Existing Probability Weighting Functions

Figures 3 and 4 exemplify differences between the functional and empirical properties of $VWD(p)$ as compared to existing weighting functions. Figure 3 shows the shape of $VWD(p)$ for different n together with parameter estimates and functional shapes of the Tversky-Kahneman, 2-parameter-Prelec, and linear-in-log-odds weighting functions when fitted to $VWD(p)$ with probability distributions of $n = 2$ to $n = 10$ that are maximally uniform.

As portrayed in the figure, $VWD(p)$ produces several fixed points without an alteration in parameter values: when n increases, the fixed point of the function changes. The other three functions, on the other hand, require different parameter value combinations to change the fixed point. In other words, when embedded in a distribution of different n, $VWD(p)$ allows for the same probability to be weighted differently, whereas existing probability weighting functions presume that the same probability receives the same weight regardless of n.

Equivalently, Figure 4 illustrates the impact of the entropy of the distribution that a probability p is embedded in on the shape of $VWD(p)$, together with fitted parameters and shapes of the Tversky-Kahneman, 2-parameter-Prelec, and linear-in-log-odds weighting functions. As with variation in n, $VWD(p)$, changes curvature depending on the shape of the distribution, whereas adjusting the curvature of existing weighting functions requires adjusting their parameters.

Formally, the flexibility of $VWD(p)$ as compared to other weighting functions can be demonstrated in the $n = 2$ case when $VWD(p)$ is compared to linear-in-log-odds. When $n = 2$, $VWD(p)$ reduces to

$$VWD_{n=2}(p) = \frac{p^{1-\frac{H_{dist}}{H_{max}}}}{p^{1-\frac{H_{dist}}{H_{max}}} + (1-p)^{1-\frac{H_{dist}}{H_{max}}}}, \qquad (5)$$

which is equivalent to the linear-in-log-odds weighting function without the elevation parameter β and with the curvature parameter α set to $1 - H_{dist}/H_{max}$. Because H_{dist} is influenced by p and H_{max} is fixed for n (for $n = 2$, $H_{max} = 1$), $1 - H_{dist}/H_{max}$ changes when p changes. α, on the other hand, is a constant.

To summarize, unlike existing weighting functions, *VWD*(*p*) predicts that the weight given to a probability *p* depends on the probability distribution (gamble) that *p* is embedded in in two ways. Firstly, whether *p* is under- or overweighted is determined by whether it is above or below the fixed point, which varies depending on the number of probabilities *n* in the distribution. Secondly, the amount of over- or underweighting of *p* depends on the relative entropy of the distribution, which is governed by its shape.

Hence, if both *VWD*(*p*) and existing probability weighting functions were fitted to data, *VWD*(*p*) should be able to capture qualitative patterns in the variation of the weight given to the same probability *p* across different *n* and distributional shapes, whereas existing weighting functions would predict the same weight for *p* regardless of *n* and distributional shape. Quantitatively, when fitted to subsets of data with different *n*, the parameters of existing probability weighting functions should change following the trends depicted in Figure 3, and fitting to subsets of data with different distributional shapes (amounts of entropy) should induce changes in parameter fits comparable to the ones presented in Figure 4. When fitted to aggregate data of gambles with different *n* and distributional shapes, *VWD*(*p*) should fit the data significantly better than existing probability weighting functions because it can account for their impact without an adjustment in parameter values.

Plausibility of *VWD*(*p*) in Light of Earlier Studies

Earlier studies in decision making under risk have found average α and β estimates ranging from 0.53 to 1 and 1 to 1.18, respectively, for the 2-parameter-Prelec weighting function (Cavagnaro et al., 2013; Stott, 2006), from 0.44 to 1.59 and 0.21 to 0.88, respectively, for the linear-in-log-odds function (Cavagnaro et al., 2013; Gonzalez & Wu, 1999; Stott, 2006), and γ averages in the range of 0.50 to 0.96 for the Tversky-Kahneman weighting function (Stott, 2006). Based on previous findings, Prelec (1998) presumed α to average at around 0.65 and β at 1. Tversky and Kahneman (1992) estimated the average fixed point to be located at around 0.34 for gains and 0.38 for losses, whereas Tversky and Fox (1995) found the fixed point to average at around 0.30, and Gonzalez & Wu (1999) at around 0.39. Given that these studies use gambles with $n = 2$ or $n = 3$ outcomes, the fitted parameter values are consistent with the estimates depicted in Figure 3 and Figure 4. However, little data exists on choices made between gambles that involve distributions of $n > 3$ outcomes or that compare across gambles with different distributional shapes.

In a slightly different area of study (decisions made based on experience), the hypothesis that the fixed point of *w*(*p*) should be at 1/*n* has been suggested by Fox and Rottenstreich (2003) and See et al. (2006), who propose that this is due to a guessing 1/*n* heuristic. In a proportion estimation task with $n = 2$ and $n = 4$ groups of colored dots, Zhang and Maloney (2012) find support for the 1/*n* fixed point hypothesis but not for the 1/*n* guessing heuristic and suggest that hedging towards 1/*n* is driven by some other cognitive process that is so far unknown. Their results are in line with results by Attneave (1953), who finds that the estimated frequency of letters in the English language ($n = 26$) is regressed towards 0.044, which is far from 0.50 (the conventional fixed point presumed in proportion estimation) and close to 1/26.

More generally, Zhang and Maloney (2012) suggest that probability and frequency related information is mentally represented as log odds, and present data of similar kinds of distortion functions that arise in a multitude of research areas across perception and cognition. A log odds representation, however, has to be complemented with additional assumptions to explain the 1/*n* phenomena. A task-general process account that resembles the framework proposed here is the assumption of a 1/*n* prior that biases probability or frequency estimates (e.g. Martins, 2006). However, in the context of decision making under risk, probabilities are not estimated, but given. Therefore, an ad hoc explanation has to be derived for how a prior can bias *adopted* (not estimated) numerical probabilities. The present approach explains the same results without making additional assumptions.

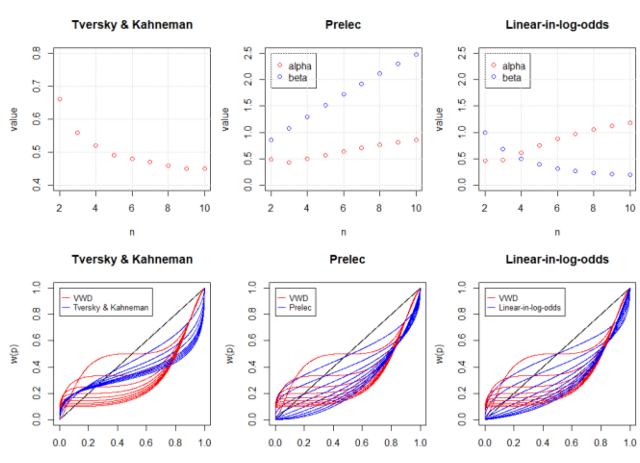

Figure 3: Upper panels: least squares estimates of Tversky-Kahneman γ (left panel, y-axis), 2-parameter-Prelec α and β (middle panel, y-axis), and linear-in-log-odds α and β (right panel, y-axis) fitted to *VWD*(*p*) for *n* ranging from 2 to 10 (x-axis). Lower panels: Fitted Tversky-Kahneman, 2-parameter-Prelec, and linear-in-log-odds functions (blue curves) on *VWD*(*p*) (red curves).

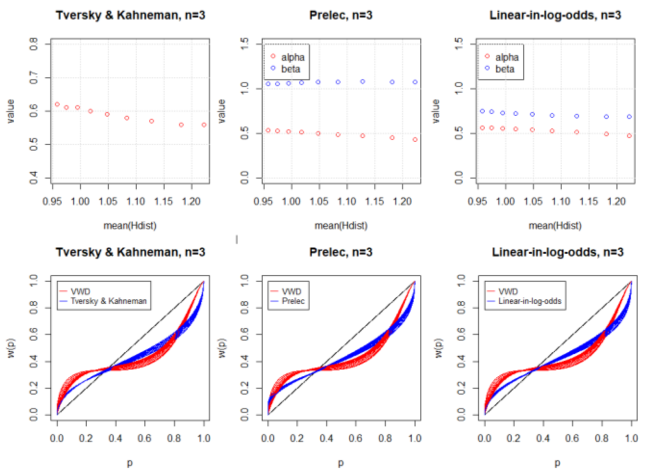

Figure 4: Upper panels: least squares estimates of Tversky-Kahneman γ (left panel, y-axis), 2-parameter-Prelec α and β (middle panel, y-axis), and linear-in-log-odds α and β (right panel, y-axis) fitted to *VWD*(*p*) for *n* = 3 probability distributions with different levels of entropy (x-axis). Lower panels: Fitted Tversky-Kahneman, 2-parameter-Prelec, and linear-in-log-odds functions (blue curves) on *VWD*(*p*) (red curves).

Contributions of the Proposed Approach

Probability Weighting and Choices under Risk

Novel Empirical Predictions If the approach presented in this paper would hold true, the weight given to a probability *p* should change when *n* is increased and the entropy of the distribution is altered as depicted in Figure 3 and Figure 4. Consequently, when fitted across gambles with different *n* and distributional shapes, *VWD*(*p*) should both capture qualitative trends and fit the data better than existing probability weighting functions. However, because individuals might differ in the amount of attention given to the number and magnitude of probabilities, *VWD*(*p*) would most likely need to be complemented with individual difference parameters when comparing to existing weighting functions.

Psychological Explanation of Existing Findings In the choice literature, Kahneman (2003) and others have suggested that the shapes of probability weighting and value (utility) functions reflect the principle of *diminishing sensitivity*: probabilities and outcomes close to a reference point (0 or 1 for probabilities, status quo for wealth) are more discriminable than probabilities and outcomes further from the reference point. The present approach explains the location of these reference points with sensitivity to uncertainty (probabilities of 0 and 1, as well as status quo, are points of minimum entropy) and predicts that the point of least discriminability lies at maximum entropy.

The proposal advanced here differs from the psychological interpretation of α and β given by Gonzalez and Wu (1999), who suggest that the α parameter reflects discriminability, whereas the β parameter reflects attractiveness of gambling. Because ibid. used certainty equivalents for two-outcome gambles to assess probability weighting, it is possible that their estimates of β reflect also the attractiveness of the gamble. However, in general, the present approach would suggest that both α and β reflect the uncertainty of the distribution that *p* is embedded in.

Combining Entropy with Expected Utility The idea of combining expected utility with information entropy is not entirely novel in the area of decision making under risk. Luce et al. (2009) suggest combining the utility derived from gambling with the expected utility of outcomes. In a related approach, Yang and Qui (2014) propose an Expected Utility-Entropy (EU-E) model that takes the form

$$EU\_E(X) = (1-\lambda) * \sum_{i=1}^{n} p_i * u(x_i) - \lambda * Hdist \qquad (6)$$

where $\boldsymbol{\lambda}$ denotes the weight given to expected utility vs. entropy.

These approaches, however, presume that expected utility and entropy are evaluated separately and that the penalty associated with entropy is uninfluenced by outcome magnitude. This appears unlikely given that the uncertainty associated with positive and negative outcomes is treated differently (e.g. Kahneman & Tversky, 1979). The approach presented here, on the other hand, embeds the influence of entropy in the evaluation of expected value, which makes the theory more parsimonious and provides one resolution to the problem of combining valence with uncertainty in psychological applications of information theory (see e.g. Luce, 2003).

Other Areas of Experimental Psychology

Parsimonious Explanation of Earlier Findings In a similar fashion as *VWD*(*p*) can be shown to be an extension of the linear-in-log-odds weighting function, it can also be shown to be an extension of the log odds representation proposed by Zhang and Maloney (2012).

Zhang and Maloney's general representation of frequency and probability weighting functions is

$$log\frac{w(p)}{1-w(p)} = \gamma * log\frac{p}{1-p} + (1-\gamma) * log\frac{p_0}{1-p_0} \qquad (7)$$

where p_0 denotes the intercept (fixed point) of the weighting function and γ is the slope of the logit function, i.e. the curvature of the weighting function.

Expressed in the same format, the linear-in-log-odds weighting function is

$$log\frac{w(p)}{1-w(p)} = \gamma * log\frac{p}{1-p} + \tau \qquad (8)$$

where γ is the curvature of the weighting function (α in our context) and $\beta = e^{\tau}$.

An equivalent expression to (7) is

$$log\,\frac{w(p)}{1-w(p)} = log\,\frac{p^{\gamma}}{(1-p)^{\gamma}} + log\,\frac{p_0^{1-\gamma}}{(1-p_0)^{1-\gamma}} \qquad (9)$$

and the *VWD*(*p*) weighting function can be expressed as

$$log\,\frac{w(p_i)}{1-w(p_i)} = log\,\frac{p_i^{\gamma}}{\sum_{j=1}^{n} p_j^{\gamma}} \qquad (10)$$

where $i \neq j$, and $\gamma = 1 - H_{dist}/H_{max}$. Comparing (9) and (10) shows that *VWD*(*p*) generalizes the log odds function by replacing the weighted log odds of *p* (i.e. the log of the ratio of *p* to $1 - p$) with the logged ratio of (weighted) *p* to every other (weighted) probability in the same distribution. This replacement eliminates the need to add a β parameter or a fixed point p_0 to the functional definition because it sets p_0 implicitly to $1/n$. The proposed function also restricts the log odds function by setting γ to the redundancy of the distribution, due to which *VWD*(*p*) makes stronger a priori predictions.

Redundancy in Psychophysics? It is widely known that studies on Stevens's (1957) power law find different exponents for different kinds of quantities, such as for length, area, or volume. In light of the proposed framework, these exponents could reflect entropic properties of the context in which the quantity is estimated in a similar fashion as the slope of the log odds function reflects the redundancy of the choice context.

Formally, this can be illustrated using the Weber-Fechner law (a special case of Stevens's power law) defined as

$$p = k * log\,\frac{S}{S_0} \qquad (11)$$

where *p* is the perceived change in a physical stimulus, *k* is a sense-specific constant, *S* is the quantity of interest (novel stimulus), and S_0 is the reference value. This can be equivalently expressed as

$$p = log\,\frac{S^k}{S_0^k} \qquad (12)$$

which, when compared to (10), suggests that the sense-specific constant *k* could be determined by the redundancy of the stimulus environment and that each novel stimuli would be compared to every older stimuli in the evaluation context. In other words, if the number of reference stimuli were increased from one (the $n = 2$ case) to $n > 2$, the reference value would be a weighted average (or other combination) of the intensities of all stimuli (similarly as the reference point is $(\sum p)/n = 1/n$ in the present case) and the magnitude of perceived change would be altered correspondingly.

Conclusion

This paper proposed a line of research that applies information theoretic notions to decision making under risk and uncertainty, aiming to provide a psychological explanation for properties of existing probability weighting functions and to introduce testable novel predictions. Even though compatible with existing data, the predictions generated by this framework will still need to be systematically tested. If successful, these results will improve the predictive ability of models applied in the study of decision making under risk and provide a psychological explanation for so far unexplained properties of these models. Furthermore, because probability weighting functions tend to share characteristics with psychophysical weighting functions, the proposed approach is likely to have applications in these areas as well. If, as the present paper presumes, the utility associated with reducing uncertainty extends from perceptual processes to the inferential processes involved in decision making under risk, it could turn out that probability weighting functions and perceptual distortion functions share the same, information encoding based root.

Acknowledgments

I would like to thank Brad Rogers, Jerome Busemeyer, Larry Maloney, and three anonymous reviewers for comments on an earlier version of this paper. This work has benefited from feedback and discussions in the Landy Lab, Logic Seminar, Cog Lunch, Judgment and Decision Making Lab, and Information Theory Study Group.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'École Américaine. *Econometrica*, *21*, 503–546.

al-Nowaihi, A., & Dhami, S. (2010). *Probability weighting functions* (Working paper No. 10/10). Retrieved from https://www.le.ac.uk/ec/research/RePEc/lec/leecon/dp10-10.pdf

Attneave, F. (1953). Psychological probability as a function of experienced frequency. *J. Exp. Psych.*, *46*, 81–86.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychol. Review*, *61*(3), 183–193.

Bernoulli, D. (1738/1954). Exposition of a new theory on the measurement of risk. Translated from Latin to English by Dr. Lousie Sommer. *Econometrica*, *22*, 22–36.

Birnbaum, M. H. (2006). Evidence against prospect theories in gambles with positive, negative and mixed consequences. *J. Econ. Psychol.*, *27*, 737–761.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychol. Review*, *115*, 463–501.

Birnbaum, M. H., Johnson, K., & Longbottom, J.-L. (2008). Tests of cumulative prospect theory with graphical displays of probability. *JDM*, *3*, 528–546.

Bouchouicha, R., Martinsson, P., Medhin, H, & Vieider, F. M. (2017). Stake effects on ambiguity attitudes for gains and losses. *Theory Dec.*, *83*, 19–35.

Bubic, A., von Cramon, D., & Schubotz, R. (2010). Prediction, cognition, and the brain. *Front. Neurosci.*, *4*(25), 1–15.

Camerer, C. F. (1989). An experimental test of several generalized utility theories. *J. Risk Uncertain, 2*, 61–104.

Camerer, C. F. (2000). Prospect theory in the wild: Evidence from the field. In D. Kahneman & A. Tversky (Eds.), *Choices, values and frames*. New York: Cambridge University Press.

Cavagnaro, D. R., Pitt, M. A., Gonzalez, R., & Myung, J. I. (2013). Discriminating among probability weighting functions using adaptive design optimization. *J. Risk Uncertain*, *47*(3), 255–289.

Crupi, V., Nelson, J. D., Meder, B., Cevolani, G., & Tentori, K. (2018). Generalized information theory meets human cognition: Introducing a unified framework to model uncertainty and information search. *Cogn. Sci.*, *42*, 1410–1456.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Q. J. Econ.*, *75*, 643– 669.

Feldman, J. (2016). The simplicity principle in perception and cognition. *WIREs Cogn. Sci.*, *7*, 330–340.

Friston, K. (2010). The free energy principle: a unified brain theory? *Nature Rev. Neurosci.*, *11*, 127–138.

Fox, C. R., & Rottenstreich, Y. (2003). Partition priming in judgment under uncertainty. *Psych. Science*, *14*, 195–200.

Garner, W. R., & Clement, D. E. (1963). Goodness of pattern and pattern uncertainty. *J. Verbal Learn. Verbal Behav.*, *2*, 446–452.

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psych. Review*, *94*, 236–254.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cogn. Psychol.*, *38*(1), 129–166.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *Am. Psycho.*, *58*, 697–720.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *TiNS*, *27*, 712–719.

Luce, R. D. (2003). Whatever happened to information theory in psychology? *Rev. Gen. Psychol.*, *7*, 183–188.

Luce, R. D. (2008). Purity, resistance, and innocence in utility theory. *Theor. Decis.*, *64*, 109–118.

Luce, R. D., Marley, A. J., & Ng, C. T. (2009). Entropy-related measures of the utility of gambling. In S. J. Brams et al. (Eds.), *The Mathematics of Preference, Choice, and Order: Essays in Honor of Peter C. Fishburn*, Studies in Choice and Welfare.

Machina, M. J. (2008). Non-expected utility theory. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*, 2nd Edition, Macmillan: Basingstone and New York.

Martins, A. C. R. (2006). Probability biases as Bayesian inference. *JDM*, *1*, 108–117.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychol. Review*, *101*, 608–631.

Prat-Carrabin, A., & Woodford, M. (2020). Efficient coding of numbers explains decision bias and noise. bioRxiv preprint, Feb 19th, 2020.

Prelec, D. (1998). The probability weighting function. *Econometrica*, *66*(3), 497–527.

Quiggin, J. (1982). A theory of anticipated utility. *J. Econ. Behav. Organ.*, *3*, 323–343.

See, K. E., Fox, C. R., & Rottenstreich, Y. S. (2006). Between ignorance and truth: partition dependence and learning in judgment under uncertainty. *J. Exp. Psychol. Learn., Mem. Cogn.*, *32*, 1385–1402.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423 and 623–656.

Stevens, S. S. (1957). On the psychophysical law. Psych. Rev., *64*(3), 153–181.

Summerfield, C., & Tsetsos, K. (2015). Do humans make good decisions? *TiCS*, *19*(1), 27–34.

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *J. Risk Uncertain*, *32*, 101–130.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertain*, *5*(4), 297–323.

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

von Neumann, J., & Morgenstern, O. (1944/1947). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Yang, J., & Qui, W. (2014). Normalized expected utility-entropy measure of risk. *Entropy*, *16*, 3590–3604.

Zhang, H., & Maloney, L. (2012). Ubiquitous log odds: a common representation of probability and frequency distortion in perception, action, and cognition. *Front. Neurosci*., *6*, 1–14.