Affective modulation of the weighting function

Victor Møller Poulsen, Studie Nr.: 201707639

January 11, 2021

Code: Simulation, figures and full analysis pipeline available at:

https://github.com/victor-m-p/BayesianDecisionWeights

Contents

1 Description

For decisions under uncertainty (e.g. bets), decision theories (e.g. Prospect Theory) can explain the value assigned to bets by combining two functions (Rottenstreich & Hsee, 2001). A value function v transforms objective value to subjective utility, and a probability weighting function w distorts probabilities (Gonzalez & Wu, 1999; Rottenstreich & Hsee, 2001). In prospect theory (PT) and expected-utility theory (EUT) the two parameters are combined in the simplest way possible (Rottenstreich & Hsee, 2001)

$$\sum w(p_i)v(i),$$

where p stands for probability and i stands for the i^{th} gamble.

That decision making under uncertainty can be formalized as a combination of a value function v and a weight function w can be construed as either an ”as-if” or a ”process” model (Newell et al., 2015). Where the ”as-if” model simply claims that people act ”as-if” these functions are computed, and the process model claims that people actually compute these functions (or something similar). This is not a primary concern here.

In EUT the weight function w is the identity $w(p) = p$, assuming that people do not distort probabilities (Rottenstreich & Hsee, 2001). Further, in EUT the value functin v is proposed to reflect how people feel about the utility of end states. This assumes that people should take into account their current state (e.g. of wealth) when evaluating outcomes (Newell et al., 2015).

With regards to both the value function v and the weight function w, PT (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) advances theorizing, and better explains empirical results (Abdellaoui et al., 2010; Wu & Gonzalez, 1996) as compared to EUT. PT is arguably the main model of human decision making (Newell et al., 2015). PT advances theorizing with regards to the value function v by positing that losses and gains are evaluated as changes in wealth rather than with regards to end states. This means that in a monetary domain, rich and poor people should show relatively similar behaviour. This is because both will evaluate outcomes based on a neutral reference point (i.e. not changing from the status quo) (Abdellaoui et al., 2010; Newell et al., 2015). This leads us to the familiar (non-linear) S-shaped value function v proposed in PT (Kahneman & Tversky, 1979). The value function v is concave for the gains domain and convex for the loss domain. This reflects the fact that small changes in outcome are relatively overweighted close to the status quo as compared to far away from the status quo. A monetary increase from $0 - 100\$$ has greater subjective utility than an increase from $1000 - 1100\$$, although the latter increase in wealth is identical in absolute terms. The same is the case for the domain of losses, where small increments are relatively overweighted close to the status quo. Another key stylistic of the value function is that it is steeper for the loss domain than for the gains domain (Newell et al., 2015). A majority of people will not take an gamble with a $0\$$ expected value (e.g. 50% chance to lose $100\$$ and a 50% chance to win $100\$$). This is the famous loss aversion.

PT advances theorizing with regards to the weight function w by allowing a (non)-linear probability distortion (Gonzalez & Wu, 1999; Kahneman & Tversky, 1979). w is stylized as being reverse S-shaped, meaning that it is concave for low probabilities and convex for high probabilities (Gonzalez & Wu, 1999; Wu & Gonzalez, 1996). This means that people underweight changes in probability in the middle of the spectrum (e.g. $[0.2 - 0.8]$) while overweighting changes in probability close to the end-points (e.g. $[0.0 - 0.2], [0.8 - 1.0]$). These general characteristics of the weighting function are empirically well documented (Abdellaoui et al., 2010; Tversky & Kahneman, 1992; Wu & Gonzalez, 1996). Prospect theory (PT) also allows for potentially different weighting functions for gains w^+ and losses w^- (Abdellaoui et al., 2010). Generally, the probability weighting function w has received less attention than the value function v (Gonzalez & Wu, 1999).

1.1 Prior work

There is evidence to support the notion that the affect-richness of outcomes modulates the parameters of both the value function v (Hsee & Rottenstreich, 2004) and the weight functon w (Rottenstreich & Hsee, 2001).

The reverse S-shape (curvature) of the weighting function w appears to be more pronounced for high-affect than low-affect outcomes under uncertainty (Rottenstreich & Hsee, 2001). Rottenstreich and Hsee (2001) investigated how much participants were willing to pay for two coupons (worth the same) at different levels of probability. The first item was a coupon redeemable for a trip to Europe (high-affect) while the second item was a coupon redeemable for covering tuition fees (low-affect). Rottenstreich and Hsee (2001) were able to show a preference reversal in which the high-affect outcome was preferred for low probability (1%) whereas the low-affect outcome was preferred for high probability (99%). If this result is solid, it suggests that people distort probabilities more for high-affect as compared to low-affect outcomes. It also violates the probability-outcome independence that both EUT and PT hold, and suggests that this independence might not hold for across outcomes which differ in affect (Rottenstreich & Hsee, 2001).

For the gains domain, the value function v has been shown to be more concave for high-affect as opposed to low-affect outcomes (Hsee & Rottenstreich, 2004). Hsee and Rottenstreich (2004) showed this by priming participants to evaluate outcomes either based on calculation or based on feeling. Their results clearly suggest a modulatory effect of affect-richness. Together, the results suggest a consistent picture of modulation of both the value function v and the weight function w. This line of evidence has been pursued elsewhere (Mukherjee, 2010, 2011) with the idea of modeling decision making as an interaction between an affective system and a deliberative system. Theoretically, the results of Hsee and Rottenstreich (2004) and Rottenstreich and Hsee (2001) are consistent with the ”risk-as-feelings” hypothesis (Loewenstein et al., 2001) which suggests that emotional reactions to decisions under uncertainty often diverge from cognitive assessment. It seems plausible that small and large probabilities of obtaining an affect-rich outcome would elicit comparably strong mental images and thus comparably strong emotional reactions.

1.2 Focus and parameterization

In this article we focus exclusively on the weighting function w while ignoring both the value function v and the combination of the two functions. We also restrict ourselves to the gains domain. A weighting function is formally a mapping from the interval $[0, 1]$ into $[0, 1]$ (Abdellaoui et al., 2010). The probability weight function can be empirically studied independently of the value function (Wu & Gonzalez, 1996). In Rottenstreich and Hsee (2001) they suggest that future work (building on their stylized results) should estimate the affective modulation on the parameter(s) of the weighting function w. This is something that they do not do in their article. They specifically propose that the affective modulation can be estimated as an affect parameter a in the general parameterization of the weighting function w:

$$w(p) = \frac{p^{1-a}}{p^{1-a} + (1-p)^{1-a}}.$$

where $a \in [0, 1]$ and larger a values indicate greater affect and more curvature (Rottenstreich & Hsee, 2001). The issue with this one-parameter formulation is that it does not account for the fact that people generally show low *elevation*. What I mean by that is that the empirically observed weighting function w typically crosses the diagonal line given by $w(p) = p$ at around $0.3 - 0.4$ rather than at 0.5 (Abdellaoui et al., 2010; Gonzalez & Wu, 1999; Wu & Gonzalez, 1996). This can be interpreted as people generally being pessimistic (i.e. 50% probability is evaluated as being worth less than 50% of the outcome). The one-parameter formulation fixes the cross-over point at 0.5, (i.e. $w(0.5) = 0.5$), which can be seen from figure 1.

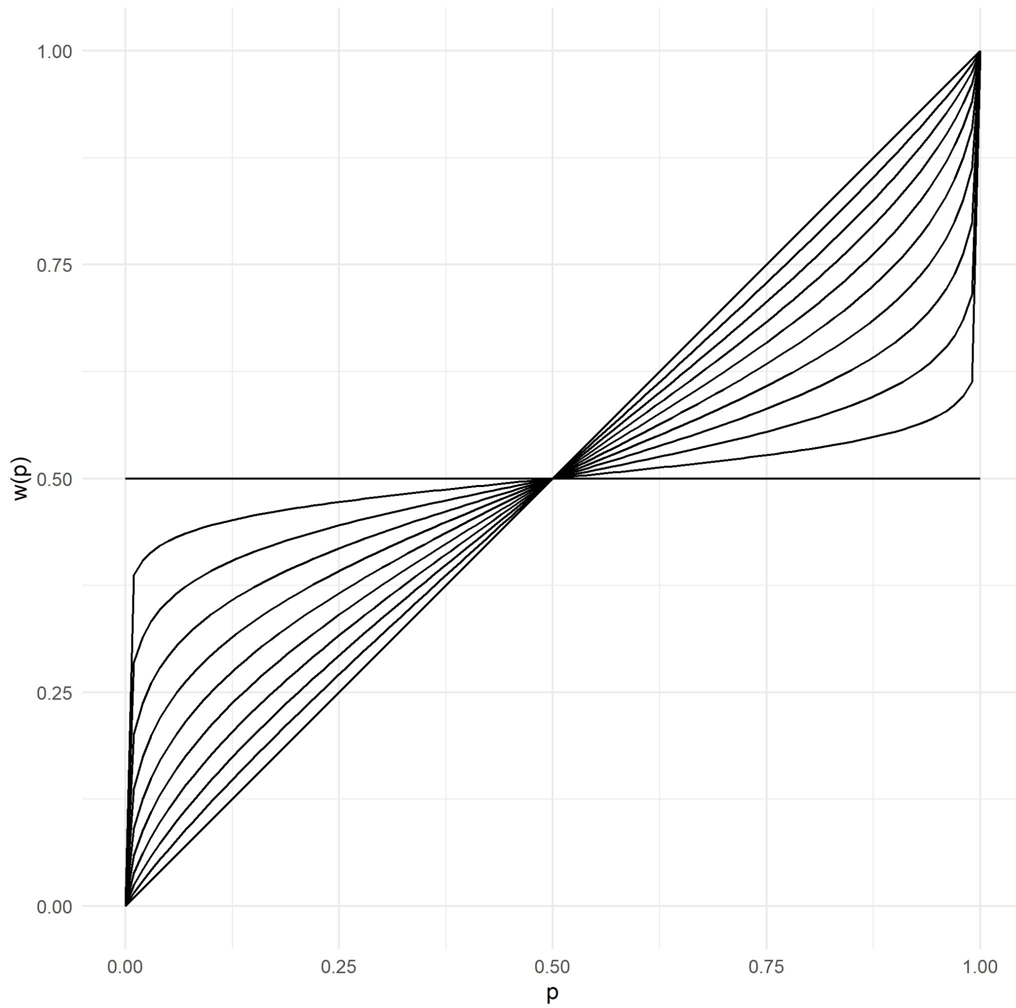

Figure 1: Data simulated from the model $w(p) = \frac{p^{1-a}}{p^{1-a}+(1-p)^{1-a}}$ with $a \in [0, 1]$ (see this file on the github). Diagonal line has $a = 0$, and the horizontal line has $a = 1$. Intermediate curves are generated for 0.2 increments of a. All values beside $a = 0$ show a probability distortion as compared to the objective probability. Note that all curves meet at $w(p) = 0.5, p = 0.5$. This is not empirically supported.

Instead of using the parameterization proposed in Rottenstreich and Hsee (2001) this paper will use the parameterization of w proposed in Gonzalez and Wu (1999).

They parameterize w with two parameters; δ and γ. Other two-parameter forms which attempt to capture elevation and curvature have been proposed, including the constant relative sensitivity (CRS) weighting functions (Abdellaoui et al., 2010).

The δ parameter will vary based on *elevation* (intercept) (Gonzalez & Wu, 1999), which here simply refers to the overall perceived attractiveness of outcomes under uncertainty.

The γ parameter will vary based on *curvature* (slope) (Gonzalez & Wu, 1999) and is what we are primarily interested in for our purposes. It follows as a direct prediction from Rottenstreich and Hsee (2001) that the curvature (γ) should be modulated by changes in the affective level of outcomes. The two parameters are logically independent, and as such they should be modeled independently and not be collapsed into one parameter (Abdellaoui et al., 2010; Gonzalez & Wu, 1999).

$w(p) = p$ for $\gamma = 1$ and $\delta = 1$ with this parameterization. Empirically, $\delta < 1$ and $\gamma < 1$ (Gonzalez & Wu, 1999). $\delta < 1$ reflects low elevation (as compared to diagonal) and $\gamma < 1$ reflects high curvature (reverse S-shape). Note that the interpretation of γ is opposite the interpretation of a as proposed in the parameterization of Rottenstreich and Hsee (2001). As such, if follows from Rottenstreich and Hsee (2001) that high-affect outcomes should result in low γ. See figure 2 for an illustration of how the δ and γ parameters independently modulate different aspects of the weighting function w.

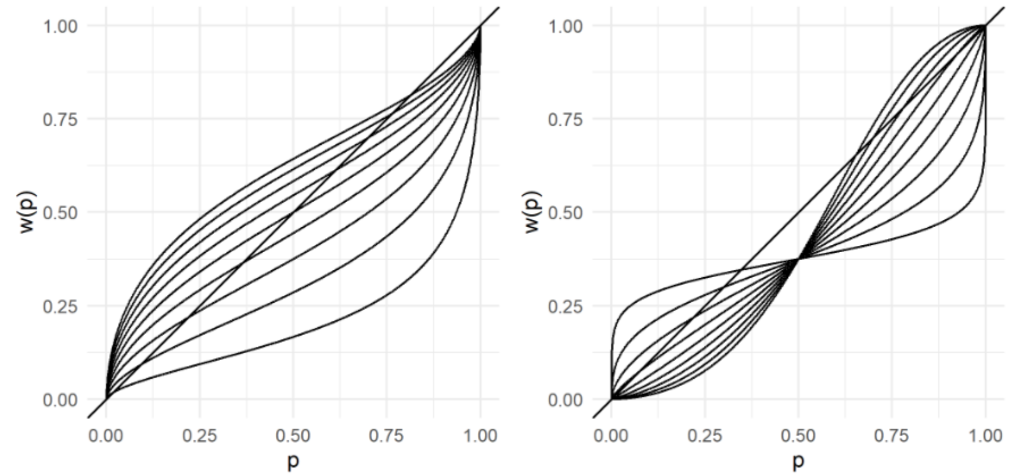

Figure 2: Data simulated from the model $w(p) = \frac{\delta \cdot p^\gamma}{\delta \cdot p^\gamma + (1-p)^\gamma}$ (see this file on the github). The figure is similar to figure 4 of Gonzalez and Wu (1999). On the left: γ fixed at 0.6 and δ varied between 0.2 and 1.8. On the right: δ fixed at 0.6 and γ varied between 0.2 and 1.8. Shows that γ controls curvature and δ controls elevation. The identity function $w(p) = p$ is achieved for $\delta = 1, \gamma = 1$.

One of the models proposed in Gonzalez and Wu (1999) is:

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

where solving for $w(p)$ and setting $\delta = \exp(\tau)$ gives us

$$w(p) = \frac{\delta \cdot p^\gamma}{\delta \cdot p^\gamma + (1-p)^\gamma}.$$

1.3 Methodology

Two studies are proposed to properly test the robustness of affect-level on the curvature (γ) of the weight function w. As such, the main goal of this study is to validate and extend the results of (Rottenstreich & Hsee, 2001) using methodology and parameterization comparable to that used in (Gonzalez & Wu, 1999).

In the first study, subjects will be asked to rate the affect-richness of 10 different items. All outcomes consist of coupons redeemable for various items, all worth $\$500$. The 10 items are designed to cover the full spectrum from affect-rich to affect-poor.

Example of expected high-affect item:
”If you won a $\$500$ coupon redeemable for a vacation abroad with a friend/partner how emotionally affected would you be?”

Example of expected low-affect item:
”If you won a $\$500$ coupon redeemable for insurance covering how emotionally affected would you be?”

For the full list of items see *Appendix A*. Participants will indicate how affect-rich each outcome is with a slider. Participants will see ”not affected at all” (left), ”somewhat affected” (middle) and ”very affected” (right). We will receive continuous ratings from 0 (affect poor) to 1 (affect rich). A mean affect rating across participants for each item will rank them from least affective to most affective. Three items are then selected: The least affective item (A), the most affective item (C) and the item in between these two extremes (B) which separate them best. We use a one-tailed t-test (directional) to compare A and B and B and C as the items (A, B, C) are by definition ranked based on mean affect rating.

In the second study, subjects will be presented with the three items (A, B, C) which have been validated for affect-richness in the prior study. All subjects rate items in all three conditions, making the study a within-subject design. The formulation around the items is that of a gamble. The formulation is the same for all items:

"You can buy a lottery ticket with an $[x]$ percent chance of winning a $\$500$ coupon redeemable for $[y]$ with a $[1 - x]$ percent chance of winning nothing. How much are you willing to pay for the lottery ticket?"

The three selected items (A, B, C) are inserted as $[y]$ and 50 different probability levels: $x = 0.01, 0.03, \ldots, 0.99$ will be inserted as $[x]$ and the negation $[1 - x]$. With all possible combinations, this means that all participants will rate the items of the 3 conditions at 50 different levels of certainty each. As in experiment 1 participants will rate with a slider. This time ranging from $\$0$ to $\$500$ as it is neither logical to assign a value below $\$0$ or above $\$500$ to any of the gambles. The approach is somewhat different from Gonzalez and Wu (1999) but ultimately we estimate the same thing that they do; participants' certainty equivalence (CE). This simply is the amount of money the gamble is worth to them. The amount of money they would accept to forego the gamble.

Note that we are not directly measuring either δ or γ. What we do measure is the dependent variable $w(p)$, and of course we also know the values of the the independent variables p and condition.

In order to infer the unmeasured parameters a bayesian (non)linear mixed effects model is proposed. The model is fitted in R (R Core Team, 2020) with the $brms$ package (Bürkner, 2018). Here we can specify the previously mentioned formula:

$$w(p) \sim \frac{\exp(\tau) \cdot p^{\gamma}}{\exp(\tau) \cdot p^{\gamma} + (1-p)^{\gamma}}.$$

It is extremely important that we specify that we want to measure $\exp(\tau)$ rather than δ as this tells the models that this parameter must be positive and thus limits the flexibility of the model in an appropriate way. δ can be inferred afterwards by taking $\exp(\tau)$. Additionally, we have to specify that the model should be nonlinear - as we believe that the weighting function is non-linear. We can further specify that we would like to estimate specifically the value for τ and γ with random intercepts (partial pooling) for participants (ID) and with item (condition) as a main effect.

$$\tau \sim 0 + item + (1|ID),$$

$$\gamma \sim 0 + item + (1|ID).$$

As mentioned, estimated τ is converted to δ by exponentiating the estimated τ value after model fitting.

Results are reported as $.66$ and $.95$ credibility intervals for the δ and γ distributions for each condition. Posterior samples are drawn from the distributions, allowing for a nice visualization of effects.

2 Hypotheses

2.1 Study 1

Hypothesis 1: As explained earlier, three outcomes are selected from the 10 investigated outcomes. The most affective (C), the least affective (A) and the question which best separate the two (B). It is hypothesized that directional t-tests between A and B and between B and C will result in significant differences. This has to be achieved before conducting study 2, as that study relies on this effect. As such, if this effect is not achieved, another study should be conducted to validate questions before proceeding with study 2. With that said however, it does appear reasonable that the 10 different questions should cover the spectrum of affect-richness pretty well (see *Appendix A*) and as such it is expected that three items which differ significantly can be extracted.

2.2 Study 2

Hypothesis 1: A directional effect is predicted for the γ parameter of the function:

$$w(p) = \frac{\delta \cdot p^\gamma}{\delta \cdot p^\gamma + (1-p)^\gamma}.$$

This follows directly from Rottenstreich and Hsee (2001). It is hypothesized that posterior credibility intervals ($.95$) for the γ parameter will not overlap between conditions, and that γ will be highest for A, lower for B and lowest for C (recall that high affect is predicted to result in high curvature, which is achieved with low

γ). This effect would replicate and seriously strengthen the results of Rottenstreich and Hsee (2001). A weaker replication would consist of credibility intervals ($.66$) showing the same effect. This would still be an interesting result, but would not be quite as convincing.

Hypothesis 2: Gonzalez and Wu (1999) report population $\gamma = 0.44$ (median). They use monetary gambles (low-affect), and as such it is hypothesized that $\gamma = 0.44$ will be within the $.95$ credibility interval of our low-affect condition (A). This would serve as a replication of the findings of Gonzalez and Wu (1999). Again, a less convincing, but interesting result would be to observe $\gamma = 0.44$ within the 0.66 credibility interval. Minimally interesting effects for γ are shown in figure 3, where γ values are $A = 0.44, B = 0.34, C = 0.24$.

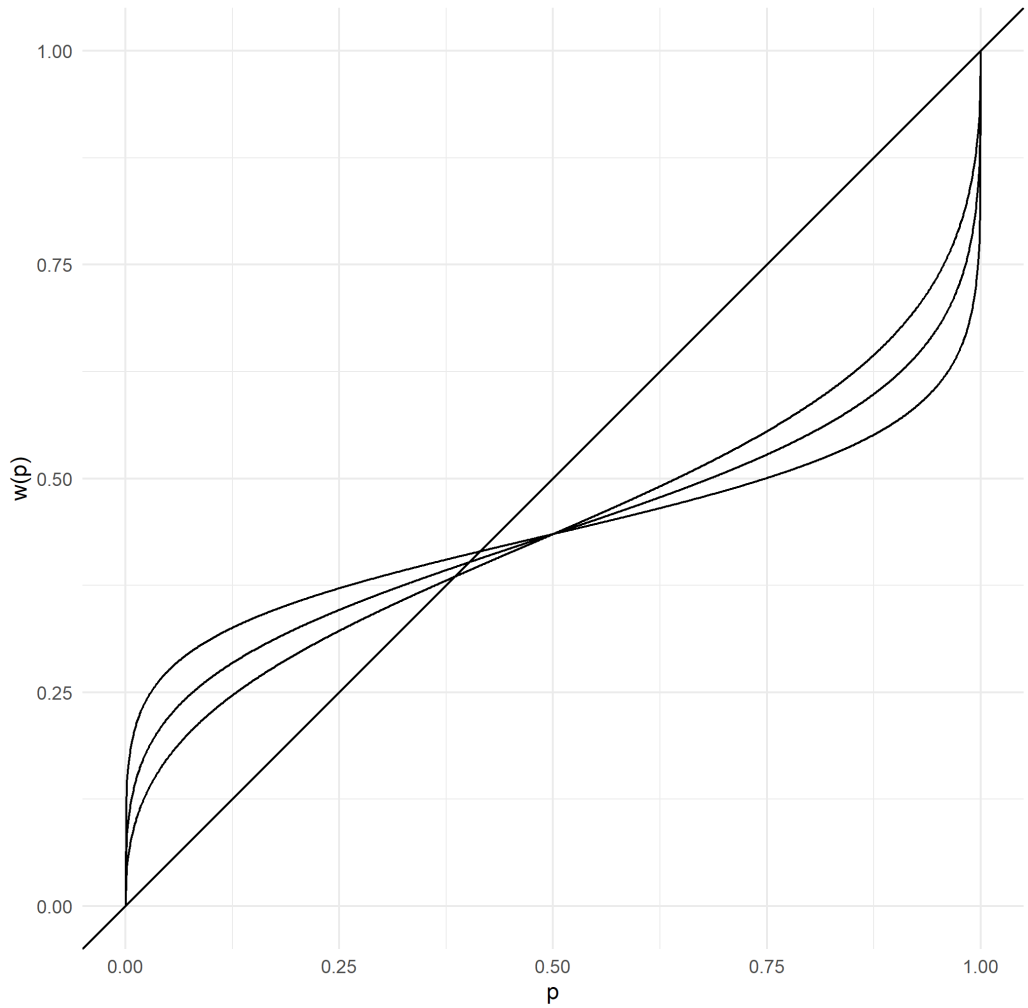

Figure 3: For code, see this file on github. Three curves shown, all with $\delta = 0.77$ as reported in Gonzalez and Wu (1999). γ levels $0.24, 0.34, 0.44$. The least curved line corresponds to $\gamma = 0.44$, as reported in Gonzalez and Wu (1999). For high-affect items γ should be lower, and as such we suggest 0.24 (for high affect) and 0.34 (for medium affect) as minimally interesting effects to detect.

Hypothesis 3: No direction of effect for the δ parameter by condition is hypothesized. The δ parameter is not of interest to the main hypothesis (replicating and extending Rottenstreich and Hsee (2001)) and is mainly included in the analysis in order to control for elevation so that the γ parameter can be estimated properly. If the three items differ in perceived overall value the δ parameter should capture this. This means that our γ distributions should still be interpretable even if the δ parameter differs by condition. The same analysis pipeline will be applied to δ as for γ (i.e. credibility intervals obtained) but as suggested, it is not clear whether an effect would be interesting. The δ parameter is expected to have a value close to .77 across conditions, which is the population median found for this parameter is Gonzalez and Wu (1999).

2.3 Simulation

In order to test the pipeline for the bayesian analysis, data simulation was conducted. Unfortunately, Gonzalez and Wu (1999) do not exactly report the values (i.e. distributional properties of τ and γ) that we need to generate data consistent with what they gathered. As such, it does not make sense to calculate power based on our simulations, and the simulation serves the sole purpose of making clear how analysis on eventual data will be conducted.

Data is generated for 50 probability levels, $p = 0.01, 0.03, \ldots, 0.99$ crossed with 3 conditions, corresponding to the actual data that will be collected. Data is generated for 30 simulated subjects (ID).

Note that standard deviations (noise) vary between γ and δ, and at different levels.

This qualitatively follows the results of Gonzalez and Wu (1999). Data is generated as a distribution of γ and δ for each condition. We generate 30 values (i) for each, corresponding to the number of participants. As we do not hypothesize that δ is modulated by condition this can simply be generated as once.

$$
\begin{aligned}
\gamma_{A_i} &\sim norm(n = 30, m = 0.44, sd = 0.1) \\
\gamma_{B_i} &\sim norm(n = 30, m = 0.34, sd = 0.1) \\
\gamma_{C_i} &\sim norm(n = 30, m = 0.24, sd = 0.1) \\
\delta_i &\sim norm(n = 90, m = 0.77, sd = 0.2)
\end{aligned} \tag{1}
$$

Based on these γ and δ values for participants per condition, we generate the final γ and δ values by adding individual noise for each probability level (j)

$$
\begin{aligned}
\gamma_{A_{ij}} &\sim norm(n = 50, m = \gamma_{A_i}, sd = 0.1) \\
\gamma_{B_{ij}} &\sim norm(n = 50, m = \gamma_{B_i}, sd = 0.1) \\
\gamma_{C_{ij}} &\sim norm(n = 50, m = \gamma_{C_i}, sd = 0.1) \\
\delta_{ij} &\sim norm(n = 150, m = \delta_i, sd = 0.3)
\end{aligned} \tag{2}
$$

As such, each condition will contain two levels of noise around a true signal, where the true population signal for γ varies by condition. For simulation code see this file on my github. The simulated data, and the best fit $w(p)$ curves are shown in figure 4. As can be seen, the simulated data shows the expected pattern, where low values of γ exhibit more curvature. The preference reversal shown in Rottenstreich and Hsee (2001) is also seen in the plot.

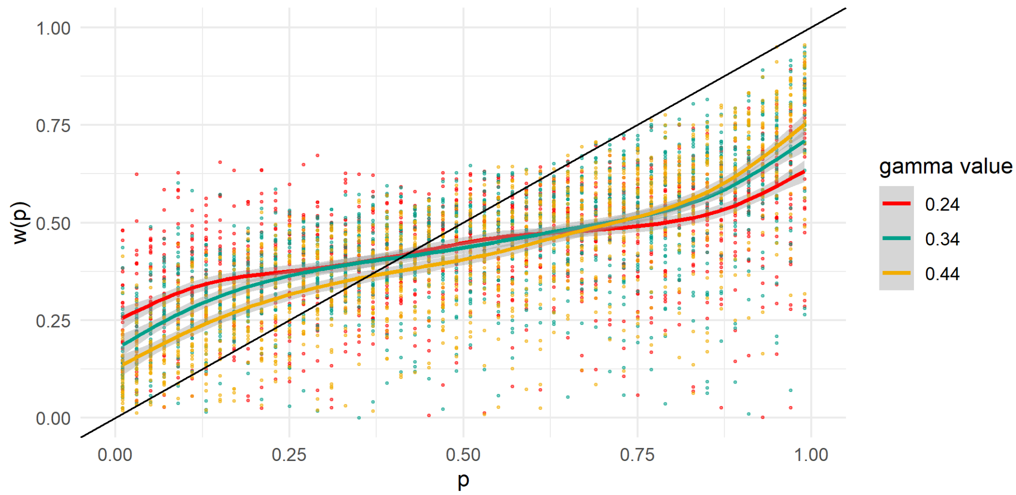

Figure 4: Plot of simulated data in three conditions. $A\,(yellow)$ has $\gamma = 0.44$, $B\,(green)$ has $\gamma = 0.34$ and $C\,(red)$ has $\gamma = 0.24$. In all conditions the true population mean of $\delta = 0.77$. Shows the preference reversal observed in Rottenstreich and Hsee (2001). Note that the yellow curve corresponds roughly to what was found in Gonzalez and Wu (1999). The population effect is a weak, but true signal, which is what we expect from the real data. Code for the figure in this file on the github

.

Next, the model described earlier is fitted to the data. See this file on the github.

Regularizing riors are specified:

$$\tau \sim normal(0, 1)$$

$$\gamma \sim normal(0.3, 0.5)$$

They reflect our knowledge of reasonable values for these parameters following both theory, and experimental results (Gonzalez & Wu, 1999). The same priors will be used for modeling the actual data. Various characteristics, such as R-hat and pp_checks (from *brms*) indicate a good model fit for the simulated data.

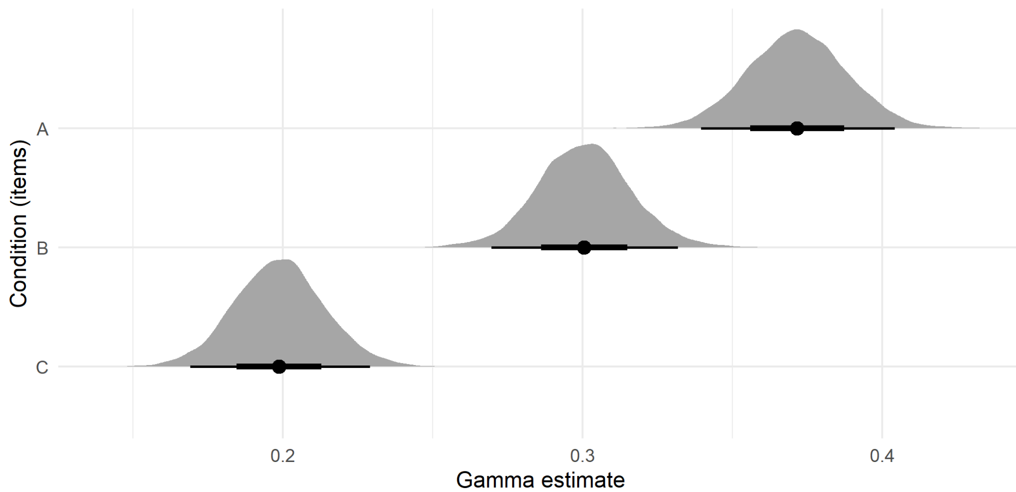

Figure 5: showing the estimated γ population distributions. The thick black line shows $.66$ credibility intervals whereas the thin black line shows 0.95 credibility intervals. Note that the conditions are ordered as expected with A having the highest γ and C having the lowest γ. Note that the .95 credibility intervals do not overlap. Code can be found in this file on the github

We extract credibility intervals with regards to γ and δ distributions for each condition (A, B, C). With the simulated data, we note that the model is capable of recovering the unobserved (unmeasured) parameters, and that for γ the $.95$ credibility intervals do not overlap between conditions (see figure 5). The estimates and credibility intervals are, $\gamma_A = 0.37, CI : [0.34, 0.40]$, $\gamma_B = 0.30, CI : [0.27, 0.33]$ and $\gamma_C = 0.20$, $CI : [0.17, 0.23]$. This slightly underestimates the true effect, which we know because we simulated the data. However, it is reasonably close. This shows that with 30 participants, and 50 probability levels crossed with 3 conditions it is possible to detect the effect that we are interested in. This of course assumes specific distributional characteristics and noise-levels that we cannot know in advance. It does however, show that the model works as intended.

3 Design Plan

Study type: Study 1 might be characterized as an observational study, since it does not really have an experimental manipulation. It resembles a survey of questions (e.g. the 10 outcomes). Study 2 is an experiment using a within-subjects design, in which all participants participate in all three conditions (A, B, C). This is important because within-subject designs have better power to detect effects than between-subject designs (Charness et al., 2012). Power is a primary concern because effects are likely to be small, and variance is likely to be high (Gonzalez & Wu, 1999). A between-subjects design is not necessary in our case, because we do not induce an effect by priming, as in e.g. Hsee and Rottenstreich (2004).

Blinding: No blinding is involved in this study.

3.1 Study Design

Study 1: All subjects will rate all 10 items (see *Appendix A*) as to the level of affect they feel with regards to them.

Study 2: All participants indicate their certainty equivalence (CE) for all certainty levels $p = 0.01, 0.03, \ldots, 0.99 (n = 50)$ and in all three conditions (A, B, C). This results in 150 observations per participant, and 50 observations per participant for each condition.

4 Sampling Plan

Existing Data: Registration prior to creation of data. Data from simulation does exist (see above & at Github).

Data collection procedures: Participants will be recruited through online channels (e.g. facebook, student groups, etc.). Participants must be at least 18 years old to participate. In the first experiments subjects will be payed 30 DKK (or $\$5$) for agreeing to participate in an approx. 10 minute online survey. In the second experiment subjects will be payed 150 DKK (or $\$25$) for agreeing to participate in an approx. 60 minute online experiment.

Sample size:

30 participants are recruited in both experiments as a minimum. Depending on funds, more than 30 participants would be preferable in especially *study 2* as this would enable us to better estimate parameters.

Sample size rationale:

Study 1: Data was simulated to estimate the approximate sample size needed to obtain a significant result from a paired t-test between A and B and between B and C. Plotting and common sense was used to arrive at best guesses for reasonable values. Three distributions were generated

$$\begin{aligned} item_A &\sim norm(0.2, 0.4) \\ item_B &\sim norm(0.5, 0.4) \\ item_C &\sim norm(0.8, 0.4) \end{aligned} \tag{3}$$

With $n = 30$ participants we have extremely high power to detect these effects with a one-tailed t-test (see this file on github). However, as this is a cheap experiment to run, and the results are critical for the second study (it is important that the three best items are used as the conditions in experiment 2) this is thought reasonable.

Study 2: Choice of sample size is naturally related to power. Typically, $.8$ power is considered reasonable (Cohen, 1992), although this is just convention. Power reflects the ability to detect a effect and is influenced by effect size and number of participants. Unfortunately, a power simulation was not possible to carry out since reasonable estimates for the distributions (and effect sizes) are not present. Additionally, there is no null hypothesis which needs a significant/not-significant label (as in a frequentist framework). Clearly, we need power to detect effects that are interesting, but several gradations of evidence are interesting. For instance, if $.95$ credibility intervals do overlap between conditions, then the somewhat weaker result that $.66$ credibility intervals do not overlap is still interesting. The simulation presented earlier likely underestimates individual variation, which will lessen power. The best comparison that we have is Gonzalez and Wu (1999) who estimate both parameters of the value function v as well parameters γ and δ of the weighting function w. They do so with 10 participants and collect data with respect to 15 gambles crossed with 11 probability levels (Gonzalez & Wu, 1999). The fact that they are able to reasonably recover the unobserved parameters (of v as well) with a sample size of only 10 participants suggests that it is not so much the number of participants, but rather the number of trials for each participant that is important. Generally there is much variation at the individual level (Abdellaoui et al., 2010; Gonzalez & Wu, 1999; Wu & Gonzalez, 1996). Based on the results of Gonzalez and Wu (1999) and on the simulation carried out, it is argued that 30 participants should probably give us reasonable power to detect minimally interesting effects.

5 Variables

5.1 Manipulated variables

Study 1: the only manipulated variable is the difference in outcome of the 10 different questions (see *Appendix A*). However, no a priori hypothesis as to which outcomes elicit high affect is necessary, and as such the experiment can be seen as close in nature to an observational survey.

Study 2: 50 levels of uncertainty are crossed with 3 conditions (different gambles). These are the two manipulated variables of study 2.

5.2 Measured variables

Study 1: The single outcome variable will be the rating of affect level. This will be measured on a scale of $0-1$. Participants will rate this using a slider (and will not see the same scale that we measure).

Study 2: The single outcome variable is the price that subjects indicate that they are willing to pay for a ticket in a lottery. This measures the certainty equivalence (CE) of participants, and can be thought of as $w(p)$. This will be measured on a scale of $0-500$ dollars using a slider. The max is 500 dollars since the lottery tickets by definition cannot be worth more than this.

5.3 Indices

No indices are used.

6 Analysis Plan

All analysis is performed in the programming language *R* (R Core Team, 2020) using *Rstudio* IDE (RStudio Team, 2020). A key package used for bayesian model fitting is *brms* (Bürkner, 2018).

Study 1: The affect ratings will be ordered based on group-level means. The three questions that best separate the are validated as being statistically significant with a paired (one-tailed) t-test.

Study 2: A bayesian generalized nonlinear mixed effects model is fit to the data using the *R* package *brms* (Bürkner, 2018). This is done to estimate the unobserved parameters δ and γ from the independent variables (1) probability level and (2) condition, and their relation to the dependent variable $w(p)$ which is the observed certainty equivalence (CE). Regularizing priors are specified for both γ and δ based on previous studies (Gonzalez & Wu, 1999). See the "simulation" section for more detail.

7 Discussion & Future Work

The two-part study presented here is important for several reasons. Firstly, it attempts to formalize an interesting finding in the field of decision making (DM). The author strongly believe that formalization is the way to make DM and psychology generally a more reproducible and predictive science. The effect that this study tries to replicate (Rottenstreich & Hsee, 2001) is very interesting. However, the study bases the claim that the probability weighting function w is modulated by affect on of a measurement of end-points (1% and 99%) only. As such, it is rather stylized; i.e. they are not able to estimate the modulation of the actual parameters of the weighting function based on their experiment. It is surprising that no-one (to my knowledge) has actually followed the suggestion of (Rottenstreich & Hsee, 2001) in extending their stylized effect. Testing it for more than two items and across enough uncertainty levels to estimate parameters of the weighting function w seems like a worthwile effort.

The present study also attempts to facilitate cumulative science more generally, by providing all code, simulation and eventual data at github. By making the code and data accessible, future experimenters can use the knowledge gained in this experiment to motivate stronger priors in subsequent experiments. This effectively pools information across studies (i.e. our posterior becomes their prior). Sadly, most of the research in this area is carried out in a frequentist framework (Gonzalez & Wu, 1999; Hsee & Rottenstreich, 2004; Rottenstreich & Hsee, 2001), and often code and data is not accessible. This makes it impossible to properly integrate previous work and thus properly build upon previous work.

If the current two-part study is successful, there are several obvious avenues of further work. Specifically, the boundaries of the modulatory effect of affect should be investigated. The proposed study only uses gambles in the gains domain (i.e. ”how much would you be willing to pay for a certain probability of winning something?”) and as such it only models w^+. To model w^-, a similar study could be run for the loss domain, still with validated high-affect and low-affect items (i.e. ”how much would you be willing to pay to avoid a certain probability of loosing something/something bad happening?”). It has been experimentally found that subjets are more sensitive to probabilities of gains (w^+) rather than losses (w^-) (Abdellaoui et al., 2010) but whether there is a modulatory effect of affect-richness is unclear.

A second avenue of further study is to broaden the analysis to include estimation of paramters of both the value function v and the weight function w. This would be interesting because the stylized effects of Rottenstreich and Hsee (2001) and Hsee and Rottenstreich (2004) could be tested together. This could validate the coherent picture of affective modulation of both v and w that these two papers seem to suggest in conjunction.

Lastly, this study assumes a specific parameterization of the probability weighting function w. Non-parametric approaches which do not assume specific functions exist, and it can be argued that these should be preferred as long as there is not agreement as to the correct functional form of the probability weighting function (Wu & Gonzalez, 1996).

References

Abdellaoui, M., l'Haridon, O., & Zank, H. (2010). Separating curvature and elevation: A parametric probability weighting function. *Journal of Risk and Uncertainty*, *41*(1), 39–65.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

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

Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology: General*, *133*(1), 23.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological bulletin*, *127*(2), 267.

Mukherjee, K. (2010). A dual system model of preferences under risk. *Psychological review*, *117*(1), 243.

Mukherjee, K. (2011). Thinking styles and risky decision-making: Further exploration of the affect–probability weighting link. *Journal of Behavioral Decision Making*, *24*(5), 443–455.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2015). *Straight choices: The psychology of decision making*. Psychology Press.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rottenstreich, Y., & Hsee, C. K. (2001). Money, kisses, and electric shocks: On the affective psychology of risk. *Psychological science*, *12*(3), 185–190.

RStudio Team. (2020). *Rstudio: Integrated development environment for r*. RStudio, PBC. Boston, MA. http://www.rstudio.com/

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

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management science*, *42*(12), 1676–1690.

8 Appendix

Appendix A

As all items in study 1 follow the same template:

”If you won a $\$500$ coupon redeemable for/at $[x]$ how emotionally affected would you be?”

The 10 proposed $[x]$ outcomes are:

- for a vacation abroad with a friend/partner.
- at a local shopping mall.
- for donation to a charity of your choice.

- for a cultural experience in your city.
- for insurance covering.
- for investing in the stock market.
- for job training.
- for $500.
- for spending on a present to someone you love.
- at your favorite restaurant.