When Is Coarseness Not a Curse? The Comparative Statics of Coarse Perception in Choice*

Sean Horan
University of Montreal and CIREQ

Paola Manzini
University of Sussex and IZA

Marco Mariotti
Queen Mary University of London

This version: 24 March 2019

Abstract

We consider an agent whose information about the objects of choice is imperfect in two respects: the true values of these objects are perceived with *error*; and the perceived values can only be discriminated *coarsely.* Reasons for coarse discrimination include limitations in sensory perception, memory function, or the technology that expert advisors use to communicate with decision-makers.

We study the effect of refining perception on the quality of decision-making. When values are perceived *without* error, finer perception is unambiguously beneficial. We show that this ceases to be true when values are perceived *with* error. As a practical implication, our results establish conditions where it is counter-productive for an expert to use a finer communication scheme with a decision-maker.

JEL codes: D01.

Keywords: stochastic choice, imprecise perception.

*We are grateful to Helios Herrera, Ian Jewitt, Mathieu Marcoux and Chris Tyson for useful discussions and advice about this project. Horan's participation was financially supported by FRQSC.

1 Introduction

An agent with finer perception may be able to discriminate between values that an agent with coarser perception cannot. We are interested in the effect of refining perception on decision-making. Specifically, *does it improve the quality of decisions?*

The pioneering work of Fechner [16] and Thurstone [36, 37] in psychology introduced the idea that decision-makers may find it difficult to discriminate between the values of different alternatives. This idea also has a long tradition in economics that dates back to Georgescu-Roegen [17], Luce [24, 25] and Quandt [32]. More recently, accumulating experimental evidence (e.g., Butler and Loomes [8, 9], Cubitt et al. [12], Permana [30]) has documented the presence of "imprecision intervals" in lottery evaluation: an agent may be confident that a lottery is worth more than $3 and less than $5, but may balk at making a more precise evaluation. As Bayrak and Hey [4, 5] have argued, this phenomenon might be the source of the observed discrepancy between willingness to pay and willingness to accept: an agent who perceives an interval of potential values might only be willing to offer the low end of the interval when acting as the buyer; and, conversely, might only be willing to accept the high end when acting as the seller.[1]

It seems intuitive that refining perception should improve decision-making. When the values of the alternatives are certain, this is indeed the case. If, for example, the agent randomises when he cannot discriminate between the values of the two alternatives, then his choices must weakly improve as his perception gets finer. In the extreme cases where the values of the alternatives are either so far apart that the coarsest perception can distinguish between them or so close that the finest perception cannot, the agent chooses the better alternative with the same probability regardless of his ability to discriminate. Between these two extremes, there are situations where finer perception causes the agent to switch from randomising to choosing the better alternative with certainty.

In the setting that we consider, the agent's inability to discriminate between values is compounded by the fact that he also perceives the values of the alternatives with noise. Like coarse perception, this type of error is well-established in psychology and economics. In the *random utility model* (Block and Marschak [6]), for instance, the perceived value of the alternatives are random variables; and the individual chooses directly according to their realisations. Just like imperfect discrimination, this model traces its origins to Fechner [16] and Thurstone [36, 37]. To take another example, decisions in the *drift diffusion model* (Ratcliff [34]) depend on a noisy process of evidence accumulation; and a choice is made only after a threshold in the difference between stimuli is exceeded.

Does finer perception remain beneficial in this noisy setting? To get some intuition

[1] Permana [30] refines the experimental methodology of Cubitt et al. [12]; and provides specific support for an *additive threshold* representation of coarse perception of the type we study in this paper.

about what might go wrong, consider the case of a referee who is asked to rank two papers, each of whose true quality is a value in $\{1, ..., 5\}$. First consider a perfectly discriminating referee who nonetheless perceives the true values with error. Suppose that the true quality of paper "A" is 5 but, with probability $^1/_{10}$, the referee wrongly perceives it as 2. In turn, paper "B" has true quality 3, which the referee always perceives correctly. Then paper "A" is correctly reported as being the better paper with probability $^9/_{10}$; and "B" is incorrectly reported as better with probability $^1/_{10}$.

Now, consider a second referee who is identical to the first in all respects except that he cannot discriminate between values (like 2 and 3) that are one unit apart. Just like the first referee, he reports that "A" is better with probability $^9/_{10}$. However, the second referee *never* reports that "B" is better. Instead, he reports that the papers are indistinguishable with probability $^1/_{10}$. Paradoxically, by heeding the advice of the referee with coarser perception, the editor is better protected from the risk of ranking the low quality paper above the high quality paper.

This example depends crucially on the interaction of random values *and* coarse perception. Decision-makers face these two different sources of error in a variety of choice environments. Imperfect memory is a source of coarse perception when indications of value come from past experiences. While a disastrous experience is almost certain to be remembered less fondly than an ideal one, it is easy for memory to blur the relative merits of two less extreme experiences. The key point is that the experiences themselves are only noisy indicators of value. Suppose, for instance, that memory follows the *peak-end rule* (Kahneman et al. [23]), where an experience is judged based on how it felt at the most intense point (the peak) and at the end of the experience. In that case, random factors, like the weather, can easily put a "shine" on a mediocre vacation; or hang a "cloud" over a more enjoyable vacation.

Expertise and information can interact to create another source of difference in perceptual coarseness with noisy values. For instance, an ill-informed voter might find it difficult to distinguish between candidates (e.g., thinking that politicians are "all the same") while an informed voter might be able to spot crucial differences. In turn, a politician's value is generally signalled imperfectly (e.g., recent events cast a disproportionately positive light on a politician whose qualities are specifically suited to those events).

Finally, note that the two sources of error need not even occur within the same agent. Consider a decision-maker (such as a policy maker, an investor, a journal editor or a juror) who relies on advice from experts. Since the experts themselves rely on scientific evidence or technical knowledge (e.g., about climate change) that is objectively uncertain, their evaluations are noisy. In addition, the experts can often make fine distinctions that they cannot adequately convey to the decision-maker (through reports or classifications of the options). As a result, the relative values of alternatives implied by expert opinions are perceived coarsely by the decision-maker.

We propose a model to capture decision-making in situations like these. We model perceptual capacity as a numerical discrimination threshold *à la* Luce [24]; and derive comparative statics for *marginal* changes in perceptual capacity. While this infinitesimal approach does not change the basic logic of "beneficial coarseness" captured in the discrete example discussed above, it simplifies the analysis considerably.

In broad strokes, our main results may be summarized as follows:

1. In general, refining perception may harm. In fact, it is only unambiguously beneficial when one alternative is superior to the other in a strong distributional sense.

2. Under some natural restrictions, whether finer perception is beneficial only depends on simple statistics (like the mean, median and mode) of the value distributions.

3. Finally, there are circumscribed but economically relevant circumstances (e.g., symmetric or identically distributed errors in values) where finer perception is beneficial.

2 Overview

2.1 The model

An agent faces the choice between two alternatives i and j, whose perceived values are represented by random variables u_i and u_j. The interpretation is that the variability represents *errors in perception of the true values* rather than taste shocks.[2]

Our analysis remains largely agnostic about the nature of the agent's errors (i.e., the relationship between the true unknown value of an alternative and the perceived values). This will allow us to highlight the rather specific nature of some parametric restrictions that are used in the literature (such as Gaussian or i.i.d. errors). We focus on two plausible scenarios, which are formalised below. In one scenario, the better alternative is *more likely to be chosen* when the agent is "standard" in the sense that he has perfect perception $\tau = 0$. This is the classical view that long run choice frequencies reflect preference (see e.g., Quandt [32]). In the other scenario, which we judge to be equally plausible, the better alternative provides the *higher expected value*.

Beyond his faulty perception of true values, the agent that we consider can only discriminate between realisations of u_i and u_j that are sufficiently far apart: the agent perceives a larger value as such only when it exceeds the lower value by a fixed threshold $\tau \geq 0$. We interpret the level of τ as a measure of the agent's perceptual *coarseness*.

[2]In Section 6.3, we show that the second interpretation dramatically changes the analysis.

When the agent can discriminate between the two values, he chooses the alternative with the higher realised value. Otherwise, he randomises uniformly between the two alternatives.[3] When the level of coarseness is τ, the probability of choosing i is given by

$$\begin{aligned} p(i,\tau) :=& \Pr[u_i > u_j + \tau] + \frac{1}{2}\Pr[\tau \geq |u_i - u_j|] \\ =& \frac{1}{2} + \frac{(\Pr[u_i > u_j + \tau] - \Pr[u_j > u_i + \tau])}{2}. \end{aligned} \tag{1}$$

In the sequel, we identify the *quality* of a decision for a given level of coarseness τ with the probability of choosing the "better" alternative. Thus, $p(i,\tau)$ measures quality when i is better; and, conversely, $p(j,\tau) = 1 - p(i,\tau)$ measures quality when j is better.

Formally, our approach grafts a random utility structure onto Luce's [24] deterministic semiorder model. As such, we maintain the assumption—central to the random utility and the drift diffusion models—that the agent chooses the alternative with the highest utility realisation. This differs from the approach recently taken by Natenzon [28], where the agent instead treats the utility realisations as signals used to update a prior.[4]

In the most general case that we consider, we allow for any (continuous) distributions of values and any pattern of correlation between the value distributions. Let F denote the joint cumulative distribution function (cdf) of the values $u_i, u_j \in \mathbb{R}$ so that $F(w,z) := \Pr[u_i \leq w, u_j \leq z]$ for all $w, z \in \mathbb{R}$. For simplicity, we assume that there exists a corresponding joint density f unless otherwise specified.[5]

For the corresponding *value difference* of the random variables $u_i - u_j$, let $f_{u_i-u_j}$ denote the density and $F_{u_i-u_j}$ denote the cdf, so that (1) can also be written as

$$p(i,\tau) = \frac{1}{2} + \frac{1}{2}\left[F_{u_j-u_i}(\tau) - F_{u_i-u_j}(\tau)\right]. \tag{2}$$

To obtain an explicit formula for the density $f_{u_i-u_j}$, note that the equality $u_i - u_j = x$ represents the event consisting of all instances where the value of j realises at $\hat{z}$ (resp. $\hat{z} - x$) and the value of i realises at $\hat{z} + x$ (resp. $\hat{z}$). Integrating over these events gives

$$f_{u_i-u_j}(x) = \int_{\mathbb{R}} f(z+x, z)\, dz = \int_{\mathbb{R}} f(z, z-x)\, dz \quad \text{for all } x \in \mathbb{R}. \tag{3}$$

We now formalise the two notions of betterness discussed above.

[3] As explained in footnote 26, our analysis easily extends to other tie-breaking rules. Our approach also covers some other decision procedures. One example is where the agent allocates the residual probability using an exogenous heuristic.

[4] Our model is also in line with the classical statistics literature on "tied comparisons" in judgement (e.g., Glenn and David [19], Greenberg [20] and Rao and Kupper [33]).The difference is that these papers focus on estimation and hypothesis testing rather than comparative statics (an issue which, to the best of our knowledge, has not been dealt with before our work).

[5] Since much of our analysis focuses on unimodal distributions, this assumption is not overly restrictive. Such distributions (are absolutely continuous and) have a density at all points except possibly the mode.

Definition 1. For alternatives i and j with random values u_i and u_j:

(i) i is *median-better* than j if $p(i,0) > \frac{1}{2}$ or, equivalently, $med[u_i - u_j] > 0$ (where $med[u_i - u_j]$ solves $\int_{-\infty}^{med[u_i-u_j]} f_{u_i-u_j}(z)\,dz = \frac{1}{2} = \int_{med[u_i-u_j]}^{\infty} f_{u_i-u_j}(z)\,dz$).

(ii) i is *mean-better* than j if $\mathbb{E}[u_i - u_j] > 0$ or, equivalently, $\int_{\mathbb{R}} z f_{u_i-u_j}(z)\,dz > 0$.[6]

(iii) i is *better* than j if it is both median- and mean-better.

By substituting weak inequalities in (i)-(iii), one obtains weak analogs of these notions.

It is worth emphasizing that, in principle, median-betterness is directly observable from behavior. Provided that one can identify circumstances where the agent has perfect discrimination, one can then use the choice frequency of alternative i to approximate the choice probability $p(i,0)$. In contrast, mean-betterness is based on global features of the value distributions that are not directly (or at least not easily) observable from choice.

2.2 Some examples

To illustrate that the quality of decision-making need not increase as perception gets finer, we first consider a simple example where the probability is concentrated at two points:

Example 1. ***(Discrete distribution)*** Suppose that the perceived value pair $u = (u_i, u_j)$ realises at $(10, 1)$ with probability $3/4$, and at $(1, 2)$ with probability $1/4$. In this case, i is the better alternative.[7] When $\tau = 1$, the worse alternative j is chosen with probability $1/2 \Pr[1 \geq |u_i - u_j|] = 1/2 \times 1/4 = 1/8$. When the level of coarseness decreases to $\tau = 1 - \varepsilon$ for arbitrarily small $\varepsilon > 0$, the probability of choosing j increases to $\Pr[u_2 > u_1 + 1 - \varepsilon] = 1/4$ (since j is now chosen outright when $(1, 2)$ realises). Accordingly, finer perception harms the agent when $\tau = 1$.

In this example, an increase in coarseness at $\tau = 1 - \varepsilon$ makes the value difference imperceptible in the event where the worse alternative is chosen. Since this change is too small to obscure the value difference in the event where the better alternative is chosen, the overall effect is to increase the probability of choosing the better alternative.

For ease of exposition, this example assumed the value realisations to be correlated. However, correlation plays no role in the effect. Indeed, it is easy to modify the example so that the values are independent but finer perception remains harmful.[8]

---

[6] When the mean is undefined (as it is in the case of Cauchy distributions, for instance), we say that i is *mean-better* than j if the Cauchy principal value $\lim_{x\to\infty} \int_{-x}^{x} z f_{u_i-u_j}(z)\,dz$ is strictly positive.

[7] Simple computation shows that $p(i,0) = 3/4 > 1/2$ and $\mathbb{E}[u_i - u_j] = \mathbb{E}[u_i] - \mathbb{E}[u_j] = 26/4 > 0$.

[8] To illustrate, suppose that $u_i = (10, 1)$ and $u_j = (1, 2)$ both realise independently with probabilities $(3/4, 1/4)$. Then, as in Example 1, alternative i is better since $p(i,0) = 27/32 > 1/2$ and $\mathbb{E}[u_i - u_j] = 131/16 > 0$. In addition, finer perception harms at the level $\tau = 1 - \varepsilon$ since $p(i,1) = \Pr[u_i = 10] + (\Pr[u_i = 1 \;\&\; u_j = 1] + \Pr[u_i = 1 \;\&\; u_j = 2])/2 > \Pr[u_i = 10] + \Pr[u_i = 1 \;\&\; u_j = 1]/2 = p(i, 1 - \varepsilon)$.

By no means is this the end of the story. For some error distributions that are commonly used in applications, refining perception has an *unambiguously* positive impact:

Example 2. ***(Logit errors)*** Suppose that $u_i = \hat{u}_i + \varepsilon_i$ and $u_j = \hat{u}_j + \varepsilon_j$ where $\hat{u}_i, \hat{u}_j \in \mathbb{R}$, where the random errors $\varepsilon_i, \varepsilon_j$ are i.i.d. Gumbel with location $\nu = 0$ and scale $c = 1$.[9] Then, the value difference $u_i - u_j$ is logistic with location $\nu = \hat{u}_i - \hat{u}_j$ and scale $c = 1$. (While this is well-known, we provide a derivation in Appendix A.) For a given level of coarseness τ, it then follows that i "beats" j with probability

$$\Pr\left[u_i > u_j + \tau\right] = \frac{e^{\hat{u}_i}}{e^{\hat{u}_j + \tau} + e^{\hat{u}_i}}.$$

The formula for $\Pr\left[u_j > u_i + \tau\right]$ is symmetric. From equation (2), it then follows that

$$p\left(i, \tau\right) = \frac{1}{2} + \frac{1}{2}\left(\frac{e^{\hat{u}_i}}{e^{\hat{u}_j + \tau} + e^{\hat{u}_i}} - \frac{e^{\hat{u}_j}}{e^{\hat{u}_i + \tau} + e^{\hat{u}_j}}\right).$$

Evaluating at $\tau = 0$ shows that i is the median-better alternative if and only if $\hat{u}_i > \hat{u}_j$. Since the mean of a logistic distribution corresponds to its location, the same condition describes the circumstances where i is the mean-better alternative. From the formula for $p\left(i, \tau\right)$, it follows that the marginal effect of coarsening perception is

$$\frac{\partial p\left(i, \tau\right)}{\partial \tau} = \frac{1}{2}\left(\frac{e^{\hat{u}_i + \hat{u}_j + \tau}}{\left(e^{\hat{u}_i + \tau} + e^{\hat{u}_j}\right)^2} - \frac{e^{\hat{u}_i + \hat{u}_j + \tau}}{\left(e^{\hat{u}_j + \tau} + e^{\hat{u}_i}\right)^2}\right).$$

This shows that a marginal improvement in perception has (one of) two possible effects. For an agent with coarse perception ($\hat{\tau} > 0$), it reduces the quality of decision-making:

$$\frac{\partial p\left(i, \hat{\tau}\right)}{\partial \tau} < 0 \iff \hat{u}_i > \hat{u}_j.$$

For a perfectly discriminating agent ($\hat{\tau} = 0$), an increase in coarseness has no effect:

$$\frac{\partial p\left(i, 0\right)}{\partial \tau} = 0.$$

Our third example shows how a minor change alters these conclusions dramatically.

Example 3. ***(Scaled logit errors)*** As in Example 2, suppose that $u_i = \hat{u}_i + \varepsilon_i$ and $u_j = \hat{u}_j + \varepsilon_j$ and that the random errors $\varepsilon_i, \varepsilon_j$ are i.i.d. Gumbel with location zero. The only difference is that the random error ε_i is now scaled by a factor $c > 1$.[10]

The critical change from Example 2 is the fact that $u_i - u_j$ is *no longer* logistic. While the distribution lacks a simple closed form expression, it is clear that the effect of the scaling factor $c > 1$ is to skew the logistic distribution in Example 2 towards the right.

[9] For a Gumbel with location ν and scale c, $F\left(z\right) = e^{-e^{-\frac{z-\nu}{c}}}$ and $f\left(z\right) = \frac{1}{c} e^{-e^{-\frac{z-\nu}{c}}} e^{-\frac{z-\nu}{c}}$.

[10] Scaling a random variable ε by $c \in \mathbb{R}_{++}$ gives a random variable ε' that is distributed like $c\varepsilon$. When ε is Gumbel with location zero and unit scale, it follows that the cdf of ε' is $F\left(z\right) = e^{-e^{-z/c}}$.

(In Appendix A, we derive an integral representation for the cdf, from which the choice probabilities can be computed using equation (2).)

This change has two implications for our analysis: first, it pushes the mean above the median; and, second, it fattens the right tail of the distribution relative to the left. The first change has the potential to drive a wedge between our two notions of betterness. In turn, the second creates the possibility that finer perception has the opposite effect for a very coarse agent ($\tau \to \infty$) as it does for a very fine agent ($\tau \to 0$).

To illustrate these observations, suppose that $0 = \hat{u}_i < \hat{u}_j = 1/2$ and $c = 2$. The density for this parametrization is shown in Figure 1 below. By direct calculation, it is not difficult to establish the following facts:

(i) While j is the better alternative without scaling (since the median and mean value differences $u_i - u_j$ are both $-1/2$), this is no longer true after u_i is scaled. While the median and mean both increase, $med[u_i - u_j] \approx -0.211$ remains negative while $\mathbb{E}[u_i - u_j] \approx 0.078$ becomes positive. So, alternative i is both mean-better *and* median-worse than j.

(ii) At the same time, the impact of finer perception changes sharply at the cutoff $\bar{\tau} \approx 4.244$. Below this level, finer perception decreases the probability $p(i, \tau)$ of selecting alternative i; and, above this level, finer perception has the opposite effect.

From these two observations, it follows that refining perception: (1) harms the mean-better alternative i and helps the median-better alternative j for sufficiently *small* levels of τ; and (2) has the opposite effect for sufficiently *large* levels of τ.

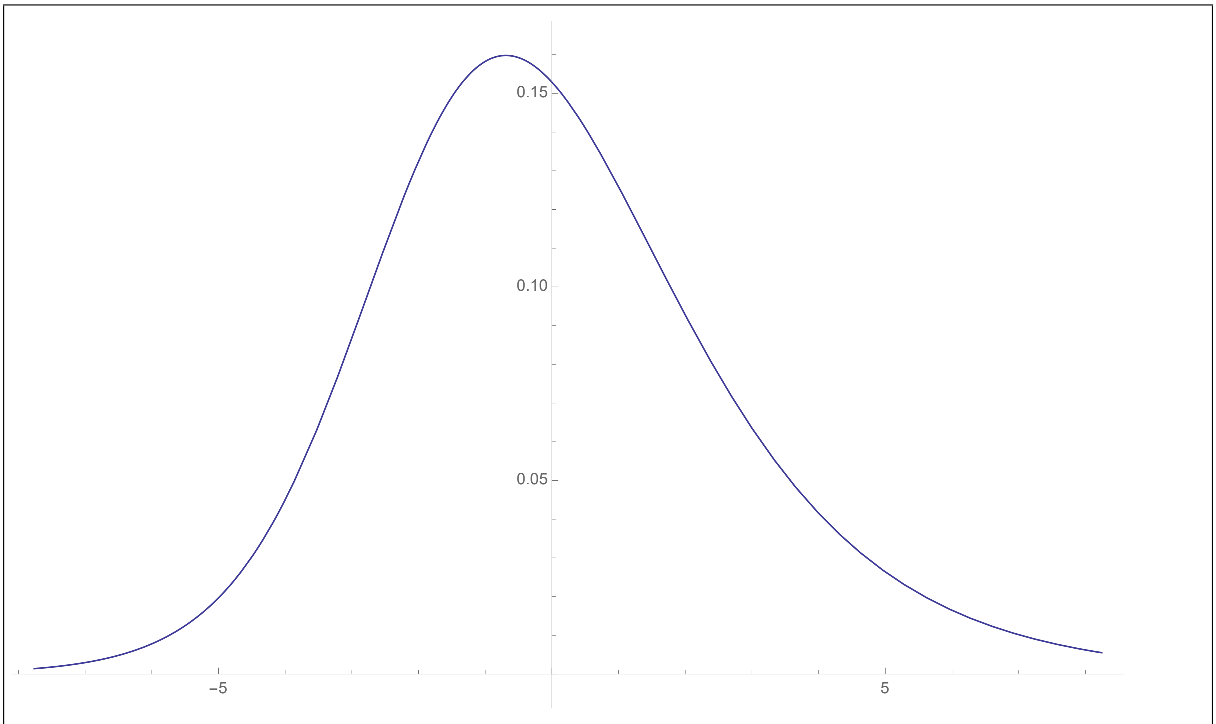

Figure 1: Plot of $f_{u_i - u_j}$ for $u_i \sim Gumbel(0, 2)$ and $u_j \sim Gumbel(1/2, 1)$

These conclusions are striking. They show that the marginal benefit (or harm) of finer perception does not necessarily bear any systematic relationship to the agent's (initially positive) level of coarseness; and may depend critically on the relevant notion of betterness.

Taken together, our three examples beg the question: *what feature of the value distributions ensure the "intuitive" effect that finer perception improves the quality of decisions?* After deriving the fundamental condition that governs the marginal impact of finer perception, we focus on value distributions that are *independent* and *unimodal*, two restrictions that hold in the most common applications of the random utility model. While a discrepancy between finer perception and quality persists even under these restrictions, they allow us to characterise the relationship in terms of primitive features of the value distributions. This, in turn, facilitates the task of doing comparative statics.

3 General analysis

In this section, we start by considering the problem in full generality. In later sections, we specialize by imposing progressively more restrictive assumptions.

Differentiating (2) and evaluating at the level of coarseness $\tau = \hat{\tau}$ immediately yields

$$\frac{\partial p\left(i, \hat{\tau}\right)}{\partial \tau} < 0 \Leftrightarrow f_{u_i - u_j}\left(-\hat{\tau}\right) < f_{u_i - u_j}\left(\hat{\tau}\right). \tag{$\star$}$$

This condition is the fundamental inequality governing the effect of finer perception in our model. It shows that the marginal impact at a given level of coarseness $\hat{\tau}$ only depends on local comparisons of the perceived value difference $u_i - u_j$. What matters are the events where $u_i - u_j$ exactly matches $\hat{\tau}$. In these *threshold events*, the probability of choosing alternative i changes by one half, either positively when the agent stops perceiving j as better (at $u_j - u_i = \hat{\tau}$); or negatively when he stops perceiving i as better (at $u_i - u_j = \hat{\tau}$). Overall, the marginal impact on the probability of choosing alternative i is half the probability difference between the threshold events.[11]

Condition ($\star$) has several notable consequences. The first is that the impact of finer perception is unrelated to the *statistical correlation* between u_i and u_j. Instead, it relates to the *cross-correlation* between u_i and u_j (in the sense of signal-processing where u_i is displaced either by a "lag" or "lead" of $\hat{\tau}$). This is the source of a systematic disconnect from the measures of quality described in Definition 1. While the effects of finer perception are driven by *local* features of the value distributions, the quality of the alternatives depend instead on *global* features of these distributions. The following calibration result puts this point in the starkest possible terms:

[11] For discrete increases in τ, the set of threshold events includes all those events where the value difference is (strictly) greater than the initial level of τ but (weakly) less than the new level of τ.

Proposition 1. ***(At any given level of coarseness, finer perception may harm unboundedly)*** *For all parameter values $\hat{\mu}, \hat{m}, \hat{\tau}, \delta > 0$, there exists a density $f_{u_i-u_j}$ that satisfies the requirements $\mathbb{E}[u_i - u_j] \geq \hat{\mu}$, $med[u_i - u_j] \geq \hat{m}$ and $\frac{\partial p(i,\hat{\tau})}{\partial \tau} \geq \delta$.*

By condition ($\star$), the sign of $\partial p\,(i, \hat{\tau})\,/\partial\tau$ is pinned down by the density of the value difference $u_i - u_j$ at exactly *two* points. So, $f_{u_i-u_j}$ is effectively unconstrained. It follows that finer perception may harm unboundedly at a given level of coarseness $\hat{\tau}$: regardless of how much better i is than j, there is some distribution of value differences for which the marginal harm exceeds a given threshold δ. (As shown in Appendix B, it is straightforward to construct such a distribution by taking an even mixture of i.i.d. distributions centred at $-\hat{\tau}$ and $2\max\{\hat{\mu}, \hat{m}\} + \hat{\tau}$, respectively.)

A second and complementary implication of condition ($\star$) is that, for a given distribution of value differences, finer perception cannot *always* harm:

Proposition 2. ***(Finer perception cannot harm at all levels of coarseness)*** *When $\hat{\tau} = 0$, finer perception has no impact (i.e., $\frac{\partial p(i,0)}{\partial \tau} = 0$). What is more, if i is either mean- or median-better than j, then $\frac{\partial p(i,\hat{\tau})}{\partial \tau} < 0$ for some level of coarseness $\hat{\tau} > 0$.*

The first statement is a direct consequence of condition ($\star$). For the second statement, suppose $\partial p\,(i, \hat{\tau})\,/\partial\tau \geq 0$ for all $\hat{\tau} > 0$. By condition ($\star$), it then follows that $f_{u_i-u_j}(-z) \geq f_{u_i-u_j}(z)$ for all $z > 0$. Integrating over this inequality gives the following:

(i) $\int_{-\infty}^{0} f_{u_i-u_j}(z)\,dz \geq \int_{0}^{\infty} f_{u_i-u_j}(z)\,dz$; and

(ii) $\int_{\mathbb{R}} z f_{u_i-u_j}(z)\,dz = \int_{0}^{\infty} z\left[f_{u_i-u_j}(z) - f_{u_i-u_j}(-z)\right]dz \leq 0$.

Observation (i) states that j is weakly median-better while (ii) implies that j is weakly mean-better. By contraposition, these observations give the desired result.

The basis for observations (i) and (ii) is yet another consequence of condition ($\star$) which, in our view, is sufficiently important to highlight separately. In particular, since $f_{u_i-u_j}(z) = f_{u_j-u_i}(-z)$, condition ($\star$) directly implies the following:

Proposition 3. ***(For finer perception to cause no harm, strong assumptions are required)*** *$\frac{\partial p(i,\hat{\tau})}{\partial \tau} \leq 0$ for almost all levels of coarseness $\hat{\tau} \geq 0$ if and only if $f_{u_i-u_j}(z) \geq f_{u_j-u_i}(z)$ for almost all value differences $z \geq 0$.*

To interpret this result, suppose that i is the better alternative. Then, for finer perception to have an unambiguously beneficial impact, u_i must dominate u_j in a strong distributional sense. Not only must u_i beat u_j "on average" so that

$$\int_{\mathbb{R}} z f_{u_i-u_j}(z)\,dz \geq \int_{\mathbb{R}} z f_{u_j-u_i}(z)\,dz$$

but u_i must also beat u_j "point-wise" so that, for almost every $z \in \mathbb{R}$,

$$z f_{u_i-u_j}(z) \geq z f_{u_j-u_i}(z)\,.$$

Clearly, this type of *point-wise dominance* is stronger than first-order stochastic dominance between the value difference distributions $u_i - u_j$ and $u_j - u_i$. In fact, it implies that the densities $f_{u_i - u_j}$ and $f_{u_j - u_i}$ cross *exactly* once at $z = 0$. (Obviously, this entails that $u_i - u_j$ first-order stochastically dominates $u_j - u_i$.) From a different angle, point-wise dominance may also be viewed as a strong form of *skewness* of $f_{u_i - u_j}$ relative to zero.[12,13]

4 Unimodal value differences

In this section, we aim to clarify the relation between quality and finer perception in a more specific setting that fits economic applications. To do so, we impose a significant but economically relevant restriction on the distribution of the perceived value differences $u_i - u_j$, namely that it is unimodal. This will allow us to express the possible disconnect between quality and finer perception in terms of simple summary statistics of the perceived value difference distribution.

Recall that a real-valued random variable X with cdf F is (*strictly*) *unimodal* if, for some value $\nu \in \mathbb{R}$, F is (strictly) convex on $(-\infty, \nu)$ and (strictly) concave on (ν, ∞). In that case, ν is a *mode* of X; and X is unimodal around ν. Since we assume that F can be associated with a density f, this is equivalent to the requirement that f is (increasing) non-decreasing on $(-\infty, \nu)$ and (decreasing) non-increasing on (ν, ∞). In that case, we say that the density f is unimodal around ν.

The preceding definition implies that a strictly unimodal distribution X has a single mode ν_X. More generally, the modes of a unimodal distribution X define a closed interval with *minimal* and *maximal* modes denoted by ν_X^{min} and ν_X^{max}. For convenience, we denote the *central mode* of a unimodal distribution by $\bar{\nu}_X := (\nu_X^{max} + \nu_X^{min})/2$.

One very large class of unimodal distributions is the class of log-concave distributions. Recall that a function $f : \mathbb{R}^n \to \mathbb{R}^n$ is *log-concave* if and only if its log-transformation is concave.[14] Many univariate distributions that are used in applications have log-concave densities (see Table 1 in Bagnoli and Bergstrom [2]). These include the normal, Gumbel, uniform, exponential, logistic, Chi-squared (with scale parameter $c \geq 2$), Gamma (with scale parameter $c \geq 1$) and Laplace (or double-exponential) distributions. A number of these distributions (including the normal, uniform, logistic and Laplace distributions) share an additional feature. They are symmetric. Recall that a function $f : \mathbb{R}^n \to \mathbb{R}^n$ is *symmetric* around the point $c \in \mathbb{R}^n$ if $f(c + y) = f(c - y)$ for all $y \in \mathbb{R}^n$.

[12] We thank Ian Jewitt for making this observation (in private communication).

[13] Macgillivray [26] considers a similar notion of skewness relative to the mean $\mathbb{E}[X]$ of a *unimodal* random variable X (see the next section for definitions). He shows that the skewness $\mathbb{E}[(X - \mathbb{E}[X])^3]$ is strictly positive if the difference $f_X(\mathbb{E}[X] + z) - f_X(\mathbb{E}[X] - z)$ changes signs *exactly* once for $z \geq 0$.

[14] In other words, $f(\lambda x + (1 - \lambda) y) \geq [f(x)]^{\lambda} [f(y)]^{1-\lambda}$ for all $\lambda \in [0, 1]$.

Our main result shows that the task of checking condition ($\star$) becomes straightforward when the distribution of $u_i - u_j$ is unimodal. Instead of performing a separate calculation for each level of coarseness $\hat{\tau}$, one may use a familiar central statistic of $u_i - u_j$ to ascertain whether condition ($\star$) holds for a range of different levels.

We start by defining a critical upper bound on the level of coarseness:

Definition 2. Suppose that the density $f_{u_i-u_j}$ is unimodal around $[\nu^{min}, \nu^{max}]$. Then, the level of coarseness $\hat{\tau} \geq 0$ is *non-confounding* if $\hat{\tau} < \max\left(\left|\nu^{min}\right|, \left|\nu^{max}\right|\right)$.

If the level of coarseness is non-confounding, then the agent is capable of perceiving which alternative is mean-better at the most extreme modal realisation of the value difference. (If the most extreme mode is $\nu > 0$, for instance, then $\hat{u}_i = \hat{u}_j + \nu > \hat{u}_j + \hat{\tau}$ and the agent perceives $\hat{u}_i > \hat{u}_j$. A similar argument applies when $\nu < 0$.)

When the distribution of the value difference is unimodal *and* symmetric, it turns out that the marginal impact of finer perception only depends on the sign of the central mode $\bar{\nu}$. In the more general case where the distribution of the value difference is unimodal but asymmetric, this is true only for non-confounding levels of coarseness.

Proposition 4. ***(The impact of finer perception depends on the sign of the central mode)*** *Suppose that the density $f_{u_i-u_j}$ is unimodal with central mode $\bar{\nu} \geq 0$. Then:*

(i) $\frac{\partial p(i,\hat{\tau})}{\partial \tau} \leq 0$ for all non-confounding levels of coarseness $\hat{\tau}$; and

(ii) if $f_{u_i-u_j}$ is also symmetric, then $\frac{\partial p(i,\hat{\tau})}{\partial \tau} \leq 0$ for all levels of coarseness $\hat{\tau}$.

What is more: if $\bar{\nu} > 0$, then there is some non-confounding $\hat{\tau} > 0$ such that $\frac{\partial p(i,\hat{\tau})}{\partial \tau} < 0$.

Proof: (i) Since $\bar{\nu} \geq 0$, $\left|\nu^{max}\right| \geq \left|\nu^{min}\right|$. Fix a non-confounding $\hat{\tau}$. Since $\hat{\tau} \in [0, \nu^{max})$ and $f_{u_i-u_j}$ is unimodal, $f_{u_i-u_j}(-\hat{\tau}) \leq f_{u_i-u_j}(\hat{\tau})$. So, $\partial p(i, \hat{\tau}) / \partial \tau \leq 0$ by condition ($\star$).

(ii) If $\hat{\tau} \in [0, \nu^{max})$, then the argument in (i) implies $\partial p(i, \hat{\tau}) / \partial \tau \leq 0$. So, suppose $\hat{\tau} \geq \nu^{max}$. By symmetry around $\bar{\nu}$, $f_{u_i-u_j}(-\hat{\tau}) = f_{u_i-u_j}(\hat{\tau} + 2\bar{\nu})$. Since $\bar{\nu} \geq 0$, $f_{u_i-u_j}(-\hat{\tau}) = f_{u_i-u_j}(\hat{\tau} + 2\bar{\nu}) \leq f_{u_i-u_j}(\hat{\tau})$ by unimodality. So, $\partial p(i, \hat{\tau}) / \partial \tau \leq 0$ by condition ($\star$).

To establish the last part of the statement, pick $\tau^* \in (0, \bar{\nu})$ if $f_{u_i-u_j}$ is strictly unimodal; and, otherwise, pick $\tau^* \in \left(\left|\nu^{min}\right|, \left|\nu^{max}\right|\right)$. In either case, it follows directly that $f_{u_i-u_j}(\tau^*) > f_{u_i-u_j}(-\tau^*)$. By condition ($\star$), this implies $\partial p(i, \tau^*) / \partial \tau < 0$. ■

Proposition 4 does not require the sign of the marginal impact to be completely determined by the sign of the central mode. (While it does require $\partial p(i, \hat{\tau}) / \partial \tau = 0$ when $\bar{\nu} = 0$, it does not rule out the possibility that $\partial p(i, \hat{\tau}) / \partial \tau = 0$ for some relevant levels of coarseness when $\bar{\nu} > 0$.) When the value difference is *strictly* unimodal however, the sign of the central mode *is* completely determinative:

Corollary 1. *Suppose that $f_{u_i-u_j}$ is strictly unimodal around $\nu > 0$. Then:*

(i) $\frac{\partial p(i,\hat{\tau})}{\partial \tau} < 0$ for all non-confounding levels of coarseness $\hat{\tau} > 0$; and

(ii) if $f_{u_i-u_j}$ is symmetric, then the same holds for all levels of coarseness $\hat{\tau} > 0$.

Much like the results in Section 3, the last two results draw a sharp distinction between the quality of an alternative and the marginal effect of finer perception. Whereas the *mean* or *median* of the value difference determines the relative quality of the two alternatives, they show that the *mode* determines the effect of finer perception.

It is worth emphasizing that unimodality imposes no restrictions on the relative ordering of the median, mean and mode (contrary to the misconception that the median must be between the other two).[15] This poses a challenge for applying Proposition 4 and Corollary 1 in practice. A second challenge relates to the fact that these two results are stated in terms of the value difference rather than the values themselves.

In the next section, we consider some restrictions on value distributions, frequently imposed in economic applications, that address both challenges. Under these restrictions, we can use the results from this section to identify circumstances where finer perception does improve the quality of decision-making and circumstances where it does not.

5 Unimodal and independent values

The results from the last section required the value difference $u_i - u_j$ to be unimodal. To clarify how these results can be used in applications, we now consider some economically relevant restrictions imposed directly on the primitive value distributions, namely that u_i and u_j are independent and unimodal. This begs the question: *what is the relationship between the two sets of conditions?*

Before Chung [10] gave a counter-example, statisticians wrongly believed that the difference of unimodal and independent random variables must be unimodal. Subsequently, a literature developed to identify conditions sufficient for the difference of unimodal and independent random variables to be unimodal. While the three main results from that literature (due to Hodges and Lehmann [21], Wintner [40], and Ibragimov [22]) are well-known in statistics, they are not so widely known in economics.

In this section, we use these three results (re-stated in Appendix C) to identify specific circumstances where finer perception is harmful and others where it is not.

To state our results, it will prove convenient to "de-mean" the value distributions, decomposing $u_k := \hat{u}_k + \varepsilon_k$ (for $k = i, j$) into a constant value $\hat{u}_k \in \mathbb{R}$ and a random variable (or *error distribution*) ε_k whose mean (or Cauchy principal value) is zero. Using these definitions, we can re-state our assumption on primitives in terms of the errors:

[15] See Abadir [1] for some examples that the mean, the median, and the mode can occur in any order.

Assumption. *The errors ε_i and ε_j are independent and unimodal.*

Except where stated, we impose this assumption for the remainder of this section.

5.1 Circumstances where finer perception is beneficial

We first identify two types of conditions where finer perception cannot harm. The first type of condition (in Propositions 5 and 6) restricts the *shape* of the errors but does not limit their *scale*. In turn, the second type of condition (in Proposition 7) limits the scale of the errors, but does not restrict their shape.

Our first result stipulates that, when the errors are identical, finer perception improves the quality of decision-making. To establish this result, we rely on Hodges and Lehmann's sufficient conditions for the unimodality of the value difference $u_i - u_j$.

Proposition 5. ***(Identical errors)*** *Suppose that the errors $\varepsilon_i, \varepsilon_j$ are identical. Then, the distribution of the value difference $u_i - u_j$ is strictly unimodal and symmetric around $\hat{u}_i - \hat{u}_j$. As a result:*

(i) alternative i is better than alternative j $\iff$ $\hat{u}_i > \hat{u}_j$ and
(ii) $\hat{u}_i > \hat{u}_j \implies \frac{\partial p(i,\hat{\tau})}{\partial \tau} < 0$ for all levels of coarseness $\hat{\tau} > 0$.

Proof: Since ε_i and ε_j are i.i.d. and unimodal, Hodges and Lehmann's theorem implies that the difference $\varepsilon_i - \varepsilon_j$ is strictly unimodal and symmetric around zero. So, the value difference $u_i - u_j$ is strictly unimodal and symmetric around (its mean, median and modal value of) $\hat{u}_i - \hat{u}_j$.[16] This establishes (i). In turn, (ii) follows from Corollary 1. ■

This result covers all of the i.i.d. specifications used to model random utility, including *logit* (Gumbel) and *probit* (normal) errors. With probit errors, finer perception must improve decision-making even when the error distributions are *not* identical. Our second result uses Wintner's sufficient conditions for the unimodality of $u_i - u_j$ to show, more generally, that the same is true for *all* symmetric error distributions.

Proposition 6. ***(Symmetric errors)*** *Suppose that the errors $\varepsilon_i, \varepsilon_j$ are both symmetric (and one of the two is strictly unimodal). Then, the distribution of the value difference $u_i - u_j$ is (strictly) unimodal and symmetric around $\hat{u}_i - \hat{u}_j$. It follows that:*

(i) alternative i is better than alternative j $\iff$ $\hat{u}_i > \hat{u}_j$ and
(ii) $\hat{u}_i \geq (>)\ \hat{u}_j \implies \frac{\partial p(i,\hat{\tau})}{\partial \tau} \leq (<)\ 0$ for all levels of coarseness $\hat{\tau} > 0$.

Proof: Since ε_i and ε_j are unimodal and symmetric, Wintner's theorem implies that their difference $\varepsilon_i - \varepsilon_j$ is unimodal and symmetric around zero. (By Theorem 2 of Appendix C,

[16] When the mean of $\varepsilon_i - \varepsilon_j$ is undefined, the argument holds verbatim for the Cauchy principal value.

$\varepsilon_i - \varepsilon_j$ is strictly unimodal if one of the errors has the same feature.) As in Proposition 5, this establishes (i). In turn, (ii) follows from Proposition 4 (Corollary 1). ■

To put Proposition 6 in context, first consider the case of *probit* errors. In this special case, the unambiguous benefit of precision follows from an "idiosyncratic" feature of normal distributions: the difference of independent normals is normal even when their scale parameters (i.e., variances) differ. In fact, the same feature holds for a much broader family of unimodal distributions, called the *stable distributions*.[17] This family is divided into classes, each indexed by a stability parameter $\alpha \in (0, 2]$.[18] Among the classes of symmetric distributions, the best known are the normal ($\alpha = 2$) and Cauchy ($\alpha = 1$). Apart from the normal distributions, every symmetric α-stable distribution has "heavy tails" (and infinite variance).[19] Analogous to the case of normal distributions, the difference of two independent α-stable distributions is α-stable.[20] Since the symmetric α-stable distributions are actually *strictly* unimodal, it follows that increased precision must be beneficial when the error distributions are independent, symmetric and α-stable.

Proposition 6 captures the symmetric α-stability cases, but is much more general since *it does not require that the errors belong to the same family* (let alone the same stability class) of distributions. It implies, for instance, that finer perception cannot harm when one of the errors follows a normal distribution while the other follows a Student t.

Our third result uses some well-known bounds that situate every mode ν_X of a unimodal distribution X relative to its median $med[X]$ and mean $\mathbb{E}[X]$ (see Corollary 4 of Basu and DasGupta [3]). These bounds depend on the variance $Var[X]$ of X:

$$\frac{(med[X] - \nu_X)^2}{3}, \quad \frac{(\mathbb{E}[X] - \nu_X)^2}{3}, \quad \text{and} \quad \frac{25\,(\mathbb{E}[X] - med[X])^2}{9} \leq Var[X]. \tag{4}$$

In our framework, the variance of the error distributions

$$Var[\varepsilon_i - \varepsilon_j] = Var[\varepsilon_i] + Var[\varepsilon_j]$$

may be interpreted as a measure of the "noise" in the agent's perception.

By combining the inequalities in (4) with Ibragimov's sufficient conditions for the unimodality of $u_i - u_j$, we can show that finer perception does not harm when the coarseness is non-confounding and the noise in the agent's perception is small relative to the mean value difference $|\hat{u}_i - \hat{u}_j|$.

[17] The unimodality of all stable distributions was first established by Yamazato [41]. See Mandelbrot [27] for economic applications of these "heavy-tailed" distributions.

[18] A stable random variable $X(\alpha; \beta, c, \nu)$ is defined by four parameters: *stability* $\alpha \in (0, 2]$; *skewness* $\beta \in [-1, 1]$; *scale* $c \in (0, \infty)$; and *location* $\nu \in (-\infty, \infty)$. When $\beta = 0$, it is symmetric around ν.

[19] Like Cauchy distributions, the mean is also undefined for distributions with $0 < \alpha < 1$.

[20] More specifically, $X(\alpha; \beta_1, c_1, 0) - X(\alpha; \beta_2, c_2, 0) = X(\alpha; \frac{\beta_1 c_1^\alpha - \beta_2 c_2^\alpha}{c_1^\alpha + c_2^\alpha}, (c_1^\alpha + c_2^\alpha)^{\frac{1}{\alpha}}, 0)$.

Proposition 7. ***(Bounded noise)*** *Suppose that one of the errors $\varepsilon_i, \varepsilon_j$ is log-concave (while the other is strictly unimodal). Then, the distribution of the value difference $u_i - u_j$ is (strictly) unimodal. Provided that $3\,(Var[\varepsilon_i] + Var[\varepsilon_j]) \leq (\hat{u}_i - \hat{u}_j)^2$, it follows that:*

(i) alternative i is better than alternative $j \iff \hat{u}_i > \hat{u}_j$ and

(ii) $\hat{u}_i \geq (>)\ \hat{u}_j \implies \frac{\partial p(i,\hat{\tau})}{\partial \tau} \leq (<)\ 0$ for all non-confounding levels $\hat{\tau} > 0$.

Proof: Since one of the errors is unimodal while the other is log-concave, Ibragimov's theorem implies that their difference $\varepsilon_i - \varepsilon_j$ is unimodal. (The corresponding statement about strict unimodality follows from Theorem 1 of Appendix C.) This establishes the (strict) unimodality of the value difference $X := u_j - u_i$.

For part (i), note that the second and third inequalities in (4) above, when combined with the restriction that $Var[u_i - u_j] = Var[\varepsilon_i] + Var[\varepsilon_j] \leq (\hat{u}_i - \hat{u}_j)^2/3$, give

$$|\mathbb{E}[X] - \nu_X| \leq |\mathbb{E}[X]| \text{ and } |\mathbb{E}[X] - med[X]| \leq \sqrt{3}\,|\mathbb{E}[X]|\,/5. \tag{5}$$

These inequalities imply that $\bar{\nu} := (\nu_X^{max} + \nu_X^{min})/2 > 0 \Leftrightarrow \mathbb{E}[X] > 0 \Leftrightarrow med[X] > 0$.

For part (ii), suppose that $\mathbb{E}[X] \geq (>)\ 0$. Then, $\bar{\nu} \geq (>)\ 0$ by the equivalence in the last paragraph; and the result follows by Proposition 4 (Corollary 1). ■

5.2 Scope for coarser perception to be beneficial

Propositions 5 to 7 identify restrictions where finer perception is beneficial. While such restrictions are common in economic applications, they are not always appropriate.

To be more specific, any difference in the way the agent perceives the two alternatives (due e.g., to differing levels of familiarity) tends to undermine the restriction in Proposition 5 (*identical errors*). In turn, the restriction in Proposition 6 (*symmetric errors*) is not likely to hold when the agent's tendency to *over*-value differs from his tendency to *under*-value (due e.g., to optimism or pessimism). Finally, even moderate noise in the agent's perceptual errors (due e.g., to a lack of familiarity with the alternatives or to their complexity) may undermine the restriction in Proposition 7 (*bounded noise*).[21]

Somewhat more concretely, consider Example 3. While the errors in this example are unimodal and independent, they violate *every* other requirement from Propositions 5-7. Clearly, they violate the shape restrictions from Propositions 5-6 (since they are Gumbel with different scale parameters). A straightforward calculation shows that these errors also violate the noise restriction from Proposition 7, given that

$$\frac{5\pi^2}{2} = 3Var[u_i - u_j] > \mathbb{E}[u_i - u_j]^2 = \left[\gamma - \frac{1}{2}\right]^2$$

[21] Burton [7] argues that because some policy interventions are complex, their outcome distributions typically exhibit heavy tails (Burton makes the case in the context of health and social care policies). In the theory of complex systems, heavy tails are considered a primary testable feature of such systems.

where $\gamma \approx 0.57722$ denotes the Euler-Mascheroni constant.[22]

Since the two notions of betterness lead to different policy prescriptions in Example 3, these observations should not come as a surprise. To drive a wedge between the two notions of betterness, the errors *must* violate the restrictions from Propositions 5-7. The only question is the potential scope for errors to create such a wedge.

Our final result in this section shows that Example 3 is not a knife-edge. It establishes that, when the difference of the errors is skewed, there exists a range of mean value pairs $(\hat{u}_i, \hat{u}_j)$ where the wedge arises. A defining feature of this range is that the mean value difference $|\hat{u}_i - \hat{u}_j|$ must remain small relative to the noise in the agent's perception. Intuitively, this shows that there is wide latitude for beneficial coarseness when the errors violate *all* of the restrictions identified in Propositions 5-7.

Proposition 8. ***(Asymmetric error differences with substantial noise)*** *Suppose that one of the errors $\varepsilon_i, \varepsilon_j$ is log-concave (while the other is strictly unimodal). Then, provided that the mean value difference $\hat{u}_i - \hat{u}_j$ is between 0 and K for some constant $K \in \mathbb{R}$ such that $K^2 \leq \frac{9}{25}\left(Var[\varepsilon_i] + Var[\varepsilon_j]\right)$, it follows that:*

(i) alternative i is mean-better and median-worse than alternative j $\iff$ $\hat{u}_i > \hat{u}_j$ and

(ii) for all non-confounding levels $\hat{\tau}_1, \hat{\tau}_2 > 0$, $\frac{\partial p(i,\hat{\tau}_1)}{\partial \tau} \leq (<)\ 0 \implies \frac{\partial p(i,\hat{\tau}_2)}{\partial \tau} \leq (<)\ 0$.

Proof: As in the proof of Proposition 7, first note that $Y := \varepsilon_i - \varepsilon_j$ is (strictly) unimodal; and that the distribution of the value difference $X := u_i - u_j$ inherits this feature.

Next, define $K := -med[Y]$ where $med[Y]$ is the median of Y. Since Y is unimodal and $\mathbb{E}[Y] = 0$ by assumption, the third inequality in (4) then implies that

$$K^2 = med[Y]^2 \leq \frac{9}{25} Var[Y] = \frac{9}{25}\left(Var[\varepsilon_i] + Var[\varepsilon_j]\right).$$

So, the constant $K \in \mathbb{R}$ satisfies the specified requirements.

For (i), fix values $\hat{u}_i, \hat{u}_j \in \mathbb{R}$ such that

$$\hat{u}_i - \hat{u}_j \in \begin{cases} (0, K) & \text{if } K \geq 0 \\ (-K, 0) & \text{otherwise;} \end{cases}$$

and note that $\hat{u}_i - \hat{u}_j > 0 \iff med[X] = med[Y] + \hat{u}_i - \hat{u}_j < med[Y] - med[Y] = 0$.

For (ii), suppose that $\hat{\tau}_1, \hat{\tau}_2 > 0$ is non-confounding. In this case, $[\nu_X^{min}, \nu_X^{max}] \neq \{0\}$. By Proposition 4 (Corollary 1), $\bar{\nu} > 0$ implies $\partial p(i, \hat{\tau}_k) / \partial \tau \leq (<)\ 0$ (for $k = 1, 2$). If $\bar{\nu} = 0$, then X is not strictly unimodal. In that case, $\partial p(i, \hat{\tau}_k) / \partial \tau = 0$ (for $k = 1, 2$). ■

[22] For a Gumbel with location $\hat{u}$ and scale c, the mean is $\hat{u} + c\gamma$ and the variance is $(c\pi)^2/6$. Since ε_i and ε_j are not mean zero, one must use $\mathbb{E}[u_i - u_j]$ instead of $(\hat{u}_i - \hat{u}_j)$ on the right side of the inequality.

6 Extensions

6.1 Beyond unimodal and independent values

The link between the mode of the value difference and the impact of finer perception remains even when the value distributions are multi-modal or dependent. (Indeed, Proposition 4 makes *no* assumptions about the individual value distributions.) The problem is that violations of unimodality or independence only further complicate the two challenges discussed at the end of Section 4. Still, it is possible to make general statements about the impact of finer perception on the quality of decision-making.

Our first result extends Proposition 6 by dispensing with independence. It shows that finer perception cannot harm for symmetric error distributions in a broad family introduced by Ghosh [18] (see also Dharmadhikari and Jogdeo [14]). Formally, a real-valued random vector $X := (X_1, X_2)$ is *linear unimodal* around the origin if, for all $a, b \in \mathbb{R}$, the linear combination $aX_1 + bX_2$ of the marginals is unimodal around zero.[23]

Proposition 9. ***(Dependence)*** *Suppose that the joint distribution of the errors* $(\varepsilon_i, \varepsilon_j)$ *is linear unimodal and symmetric around the origin. Then:*

(i) alternative i *is better than alternative* $j \iff \hat{u}_i > \hat{u}_j$ *and*
(ii) $\hat{u}_i \geq \hat{u}_j \implies \frac{\partial p(i,\hat{\tau})}{\partial \tau} \leq 0$ *for all levels of coarseness* $\hat{\tau} > 0$.

Proof: Since $(\varepsilon_i, \varepsilon_j)$ is linear unimodal around the origin, it follows that the difference $\varepsilon_i - \varepsilon_j$ is unimodal around zero. To see that $\varepsilon_i - \varepsilon_j$ is symmetric, let g denote the density of $(\varepsilon_i, \varepsilon_j)$ and $g_{\varepsilon_i - \varepsilon_j}$ the density of $\varepsilon_i - \varepsilon_j$. Then, for all $x \in \mathbb{R}$,

$$g_{\varepsilon_i-\varepsilon_j}(x) = \int_{\mathbb{R}} g(z + x, z)\, dz = \int_{\mathbb{R}} g(-z - x, -z)\, dz = \int_{\mathbb{R}} g(z - x, z)\, dz = g_{\varepsilon_j-\varepsilon_i}(x)$$

where: the first and last equalities follow by equation (3); the second by the symmetry of $(\varepsilon_i, \varepsilon_j)$ around the origin; and the third by the change of variables $z \to -z$. Since $\varepsilon_i - \varepsilon_j$ is unimodal and symmetric around zero, the result then follows from Proposition 4. ■

The error distributions covered by this result are generalizations of bi-variate normals. As noted after Proposition 6, the difference of two independent normals is normal. In fact, this is true even when the distributions are dependent, provided that they are *jointly* normal. Linear unimodal and symmetric distributions have a similar closure property: for such distributions, the difference of marginals is unimodal and symmetric around zero. Not only does this imply that finer perception cannot harm for bi-variate normal errors but it implies that the same is true for a much wider class of errors (including any symmetric error distribution that is either stable or log-concave).

[23] While Ghosh defines the class for n-dimensional random vectors, we only require two dimensions.

Generalising in a different direction, it is possible to dispense with unimodality and retain a local version of Proposition 5. As in that result, the assumption of i.i.d. errors plays a key role, ensuring that the better alternative is the one with the higher mean.

Proposition 10. ***(Multi-modality)*** *Suppose that the errors $\varepsilon_i, \varepsilon_j$ are i.i.d. with continuous densities. Then:*

(i) alternative i is better than alternative j $\iff$ $\hat{u}_i > \hat{u}_j$ and

(ii) $\hat{u}_i > \hat{u}_j \implies \frac{\partial p(i,\hat{\tau})}{\partial \tau} \leq 0$ for all levels of coarseness $\hat{\tau} \in B(|\hat{u}_i - \hat{u}_j|)$ in some open interval $B(|\hat{u}_i - \hat{u}_j|)$ around $|\hat{u}_i - \hat{u}_j|$.

Proof:[24] Let f_i and f_i denote the densities of u_i and u_j; and let $d := \hat{u}_i - \hat{u}_j$. Since ε_i and ε_j are i.i.d., $f_i(x) = f_j(x-d)$ for all $x \in \mathbb{R}$. Substituting this identity into the derivative of equation (2)with respect to τ and applying a change of variable $z \to z + d$ gives

$$\frac{\partial p(i,\hat{\tau})}{\partial \tau} = \frac{1}{2}\left[\int_{\mathbb{R}} f_j(z) f_j(z+\hat{\tau}+d)\,dz - \int_{\mathbb{R}} f_j(z+\hat{\tau}-d) f_j(z)\,dz\right]. \tag{6}$$

Since $\int_{\mathbb{R}} [f_j(z)]^2\,dz = \int_{\mathbb{R}} [f_j(z+c)]^2\,dz$ for all $c \in \mathbb{R}$, the quadratic formula yields

$$\int_{\mathbb{R}} f_j(z) f_j(z+c)\,dz = \int_{\mathbb{R}} [f_j(z)]^2\,dz - \frac{1}{2}A(c) \tag{7}$$

where $A(c) := \int_{\mathbb{R}} [f_j(z) - f_j(z+c)]^2\,dz$. By substituting (7) into (6), one obtains

$$\frac{\partial p(i,\hat{\tau})}{\partial \tau} = \frac{1}{4}\left[A(\hat{\tau}-d) - A(\hat{\tau}+d)\right].$$

As in the proof of Proposition 5, $u_i - u_j$ is symmetric around (its mean, median value of) $d = \hat{u}_i - \hat{u}_j$. Now, suppose that i is better than j (or, equivalently, that $\hat{u}_i > \hat{u}_j$). Then,

$$\frac{\partial p(i,d)}{\partial \tau} = \frac{1}{4}\left[A(0) - A(2d)\right] = -\frac{1}{4}\int_{\mathbb{R}} [f_j(z) - f_j(z+2d)]^2\,dz. \tag{8}$$

Observe that f_j is not a density if $f_j(z) = f_j(z+2d)$ for almost all $z \in \mathbb{R}$. Otherwise,

$$\int_{\mathbb{R}} f_j(z)\,dz = \sum_{k\in\mathbb{Z}} \int_{k2d}^{(k+1)2d} f_j(z)\,dz = \left[\lim_{k\to\infty} 2k\right]\left[\int_0^{2d} f_j(z)\,dz\right] \neq 1.$$

So, $f_j(z) \neq f_j(z+2d)$ for some set $A \subseteq \mathbb{R}$ of positive Lebesgue measure. This, in turn, implies $\partial p(i,d)/\partial\tau < 0$ by equation (8). By continuity of f_j, the result obtains. ■

6.2 Beyond two alternatives

It is not entirely straightforward to extend our analysis to more than two alternatives. In this section, we briefly mention two issues that complicate the task.

[24]The proof adapts standard techniques to show that cross-correlation is minimal at zero.

(1) Impact of finer perception: As the number of alternatives increases, the range of possible "ties" in the value realisations increases exponentially, which significantly complicates the task of determining the choice probabilities and, consequently, the marginal impact of refining perception.

This is already apparent with three alternatives. As in the case of two alternatives, suppose that the alternatives tied "at the top" are chosen with uniform probability. Let R_S^τ denote the probability of the event that S is the *top-set* of alternatives:

(i) for every $i \in S$, $|u_i - u_j| \leq \tau$ for all $j \in S$; and,

(ii) for every $k \notin S$, $u_k + \tau < u_j$ for some $j \in S$.

With this notation, R_{12}^τ, R_{13}^τ and R_{123}^τ reflect the events where alternative 1 is tied at the top and R_1^τ the event where it wins outright. Then, $p(1,\tau)$ can be written as

$$\begin{aligned} p(1,\tau) &= R_1^\tau + \frac{1}{2}\left[R_{12}^\tau + R_{13}^\tau\right] + \frac{1}{3}R_{123}^\tau \\ &= \frac{1}{3} + \frac{1}{3}\left[2R_1^\tau - R_2^\tau - R_3^\tau\right] + \frac{1}{6}\left[R_{12}^\tau + R_{13}^\tau - 2R_{23}^\tau\right]. \end{aligned} \tag{9}$$

The second formulation (which follows from the first by re-writing R_{123}^τ in terms of its complementary probabilities) more clearly states how $p(1,\tau)$ is affected by increases in τ. Letting $R_{S\to T}^\tau$ denote the probability of the threshold event that the top-set switches from S to T when τ increases, the marginal effect of finer perception is

$$\begin{aligned} \frac{\partial p(1,\tau)}{\partial \tau} &= \frac{1}{2}\left[R_{2\to 12}^\tau - R_{1\to 12}^\tau\right] + \frac{1}{2}\left[R_{3\to 13}^\tau - R_{1\to 13}^\tau\right] \\ &\quad + \frac{1}{6}\left[2R_{23\to 123}^\tau - R_{12\to 123}^\tau - R_{13\to 123}^\tau\right]. \end{aligned} \tag{10}$$

This shows that *three* different trade-offs determine the marginal impact of finer perception when there are three alternatives. The first two terms are direct analogs of the two-alternative case where the perceived value of one alternative changes (either positively or negatively) relative to *one* other alternative. In the last term, the perceived value of one alternative changes relative to *two* other alternatives.[25]

More generally, with n alternatives, the marginal impact of finer perception on $p(i,\tau)$ involves $2^{n-1} - 1$ different trade-offs. (For each $k = 1, ..., n-1$, there are $\binom{n-1}{k}$ trade-offs where the perceived value of one alternative changes relative to k others.)

(2) Median quality: With two alternatives, we believe that there are compelling reasons to use the median value difference as a measure of quality. However, difficulties arise in trying to generalize this measure to three or more alternatives. At a fundamental level, the issue is that the "univariate" median used for two alternatives does not necessarily identify a highest quality alternative. The following example illustrates.

[25]For the interested reader, we derive explicit formulas for equations (9) and (10) in Appendix D.

Example 4. For three alternatives, consider a random utility specification that induces the following distribution $\Pi_>$ over the ranking of value realizations

$$\Pi_> := \begin{bmatrix} \Pr[u_1 > u_2 > u_3] \\ \Pr[u_1 > u_3 > u_2] \\ \Pr[u_2 > u_1 > u_3] \\ \Pr[u_2 > u_3 > u_1] \\ \Pr[u_3 > u_1 > u_2] \\ \Pr[u_3 > u_2 > u_1] \end{bmatrix} = \begin{bmatrix} 18/64 \\ 7/64 \\ 3/64 \\ 15/64 \\ 11/64 \\ 10/64 \end{bmatrix}.$$

In this case, the (pairwise) medians induce a cyclic quality ranking. Let $p_{ij}(k, 0)$, with $k \in \{i, j\}$, denote the probability that a perfectly precise ($\tau = 0$) agent chooses k from $\{i, j\}$. Then:

$$p_{12}(1, 0) = p_{23}(2, 0) = p_{13}(3, 0) = \frac{36}{64} > \frac{28}{64} = p_{12}(2, 0) = p_{23}(3, 0) = p_{13}(1, 0).$$

We emphasize that an *independent* random utility specification is sufficient to produce the ranking distribution $\Pi_>$, specifically one where each of $u_1 = (1, 4, 7, 7)$, $u_2 = (2, 6, 6, 6)$ and $u_3 = (3, 5, 5, 8)$ realizes with uniform probability. (In the literature, independent distributions that induce such pairwise "cycles" are known as *non-transitive dice.*)

To resolve this issue, one possibility is to rely on a "multivariate" concept of median to identify the highest quality alternative. Having said this, there are multiple ways to extend the "univariate" concept and it is not clear which is the most appropriate.

A different approach would be to measure quality by the median of a "multivariate" object. One possibility is the median ordering, which is determined by ranking alternatives by the probability that they are chosen from the grand set of alternatives. When there are two alternatives, the (top-ranked alternative according to the) median ordering is the "univariate" median. For three or more alternatives, the median ordering is faithful to the idea that a perfectly precise ($\tau = 0$) agent tends to choose well. To illustrate, consider the agent in Example 4. In that case, the median ordering is $1 > 3 > 2$ since

$$\Pr[u_1 > u_2, u_3] = \frac{25}{64} > \Pr[u_3 > u_1, u_2] = \frac{21}{64} > \Pr[u_2 > u_1, u_3] = \frac{18}{64}.$$

The median ordering captures the intuition that alternative 1 is the best among the three and, as a "second-order" concern, that alternative 3 is better than alternative 2.

6.3 Taste shocks

Our analysis assumed that the random values reflect errors of perception. If the errors instead reflects taste shocks, then the realised *values* are better interpreted as welfare relevant *utilities*. In that case, it seems more appropriate to measure the quality of a decision by its expected utility.

To analyse this variation, we generalise the model by supposing that, when he cannot distinguish the two values, the agent chooses alternative i with probability $\alpha \in (0,1)$. The parameter α expresses the *bias* of the agent towards alternative j when he cannot distinguish it from alternative i.[26] Then, the agent's expected utility of choosing according to his coarse perception may be expressed as follows:

$$\mathbb{E}\left[u\left(\tau,\alpha\right)\right] := \mathbb{E}\left[u_i | u_i > u_i + \tau\right] + \mathbb{E}\left[u_j | u_j > u_i + \tau\right] + \mathbb{E}\left[\alpha u_i + \left(1-\alpha\right) u_j | \tau \geq |u_i - u_j|\right]$$
$$= \int_{\mathbb{R}} \int_{-\infty}^{z-\tau} z f\left(z,w\right) dw dz + \int_{\mathbb{R}} \int_{z+\tau}^{\infty} w f\left(z,w\right) dw dz + \int_{\mathbb{R}} \int_{z-\tau}^{z+\tau} \left(\alpha w + \left(1-\alpha\right) z\right) f\left(z,w\right) dw dz.$$

Differentiating this expression and evaluating at the level of coarseness $\hat{\tau}$ gives:

$$\frac{\partial \mathbb{E}\left[u\left(\hat{\tau},\alpha\right)\right]}{\partial \tau} = -\left[\alpha\hat{\tau} \int_{\mathbb{R}} f\left(z, z-\hat{\tau}\right) dz + \left(1-\alpha\right)\hat{\tau} \int_{\mathbb{R}} f\left(z, z+\hat{\tau}\right) dz\right].$$

Since each of the terms inside the brackets is non-negative, we conclude the following:

Proposition 11. *For all $\alpha \in (0,1)$ and every level of coarseness $\hat{\tau} > 0$:* $\frac{\partial \mathbb{E}[u(\hat{\tau},\alpha)]}{\partial \tau} \leq 0$.

This result shows that, when measured by its expected utility, the quality of the agent's decision cannot decrease with finer perception (although it may not change for certain distributions). This is true under completely general conditions that do not depend on either the joint distribution f of utilities or the size of the bias α.

7 Concluding remarks

We have studied the interplay between two distinct sources of error in judgment: the *absolute* error in perceiving the value of a given option; and the *relative* error in comparing the values of two options. We captured the distinction between these two sources of error by modelling the perceived value of each option as a random variable and the perception of the *value difference* by a "just noticeable" threshold (or *level of coarseness*) τ. Both types of imperfection (and they way we model them) are standard in psychology and economics. The novelty of our analysis lies in studying *the effect of their interaction under general hypotheses on the structure of errors.* While finer perception (i.e., a decrease in τ) is unambiguously beneficial when the agent perceives the values perfectly, matters are less clear when the agent perceives the values imperfectly. Indeed, our results characterise different classes of problems where finer perception may lead to choosing the worse alternative. Some popular parametric restrictions on the error structure (such as Gaussian or symmetric errors) hide these possibilities.

[26] A similar generalisation could be made to our main model in equation (1). It would change the right-hand side of condition ($\star$) from $f_{u_i-u_j}\left(-\hat{\tau}\right) < f_{u_i-u_j}\left(\hat{\tau}\right)$ to $\alpha f_{u_i-u_j}\left(-\hat{\tau}\right) < (1-\alpha) f_{u_i-u_j}\left(\hat{\tau}\right)$.

While our analysis is primarily theoretical, our results have practical relevance. Consider, for instance, the GRADE international standard used for evaluating scientific evidence in medical practice. This system provides a way to categorise the strength of evidence from clinical trials into different *certainty ratings*.[27] If the result of a clinical trial is a random variable, then the attribution of a certainty rating corresponds in our framework to the comparison of the trial results to a fixed benchmark; and the judgment about whether to adopt the treatment comes down to the clinician's perception of the difference between the trial results and the certainty rating.

More concretely, suppose that a physician must choose between recommending two treatments (such as an exercise regimen and a specific diet) to an overweight patient.[28] Each treatment is a random variable whose realisation reflects the aggregate weight losses of the participants in the associated medical trials. Which treatment will be judged more effective will depend on the physician's level of *confidence* (as measured by τ). According to our results, a more confident physician (with a smaller τ) may recommend the wrong treatment more often than a less confident physician (with a larger τ).

As a second illustration, consider the implications of our analysis for labor markets. To fix ideas, suppose that an applicant's performance at a job interview is a random variable; and that the prospective employer bases the hiring decision on the interview performance of the candidate relative to a benchmark (which reflects the criteria stated in the job description). In particular, suppose that the employer definitely hires an applicant whose interview performance exceeds the benchmark by a threshold difference τ and definitely does not hire one whose performance falls short by τ.

Our results show that unintended consequences can result when the threshold τ varies with observable characteristics of the applicant. To illustrate, suppose that the employer applies a smaller threshold to an applicant with a post-secondary degree (reasoning, perhaps, that such an applicant should exhibit less variance between perceived performance and true ability). Then, contrary to compensating for higher variance, our results show that an even higher proportion of errors (in either direction) could be made when hiring applicants without post-secondary education. More generally, if the level of coarseness τ (interpreted as a level of "tolerance") varies with observable characteristics like age, sex or race, it can result in unintended discrimination.

Our discussion has been carried out in the context of individual decision making, but by no means is the problem of imprecise discrimination restricted to individuals. *Collective*

[27]See Schünemann et al. [35] for full details on the GRADE framework.

[28]The example is based on Clark [11], a meta study on the efficacy of different training/dietary regimes for weight loss. This meta study reports differences in effectiveness with reference to various metrics, from body mass to fat free mass and blood levels of certain hormones. Its tables and figures summarise the difference in distributions of the relevant variables across the individual studies considered.

decision-makers also face a host of difficulties that lead to coarseness. Policy makers must balance the conflicting values of individuals in society; and the individuals themselves might only perceive or communicate their values coarsely. Danan et al. [13], for example, study these issues in the context of social decisions, focussing on imprecise *beliefs* and on the Pareto principle as a means of singling out decisions that are robust to imprecise beliefs. In future research, it would be interesting to explore the specific consequences of our analysis in such collective settings.

To close, it is worth noting that our analysis assumed a simple, one-parameter, threshold structure. We took this approach because it allowed us to study the comparative statics of imprecise judgment in a sharp way with as few "moving parts" as possible. Having said this, it would be interesting to see how our analysis could be extended to more structured situations, such as when: (a) there is a threshold τ_{ij} for alternatives i and j that depends on the pair being compared in the form $\tau_{ij} = \tau_i + \tau_j$ (i.e., each alternative is has its own "inborn" level of coarseness and the coarseness in any comparison is the total coarseness of the alternatives being compared); (b) there are two different thresholds, τ_{ij} and τ_{ji}, such that i is chosen over j when $u_i - u_j > \tau_{ij}$ and j is chosen over i when $u_j - u_j > \tau_{ji}$.[29] We leave this investigation to further research.

References

[1] Abadir, Karim M. (2005) "The Mean-Median-Mode Inequality: Counterexamples." *Econometric Theory* 21(2): 477-482.

[2] Bagnoli, Mark, and Ted Bergstrom (2005) "Log-concave Probability and its Applications." *Economic Theory* 26(2): 445-469.

[3] Basu, Sanjib, and Anirban DasGupta (1997) "The Mean, Median, and Mode of Unimodal Distributions: a Characterization." *Theory of Probability and its Applications* 41(2): 210-223.

[4] Bayrak, Oben K. and John D. Hey (2016) "Is there a valuation gap? The case of interval valuations." *Economics Bulletin* 36(1): 218-236

[5] Bayrak, Oben K. and John D. Hey (2017) "Expected utility theory with imprecise probability perception: explaining preference reversals." *Applied Economics Letters* 24(13): 906-910.

[29]Both of these ideas have appeared in the statistical estimation literature on tied comparisons (that we mentioned in footnote 4). Another recent idea regarding the parametric modelling of coarseness is proposed by Tyson [38]. In his model, the probability that a given utility difference is perceived is assumed to decrease exponentially with the size of the difference.

[6] Block, H., and J. Marschak (1960) "Random Orderings and Stochastic Theories of Responses" in Olkin, Gurye, Hoeffding, Madow, and Mann (eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling.* Stanford University Press.

[7] Burton, C. (2012) "Heavy Tailed Distributions of Effect Sizes in Systematic Reviews of Complex Interventions." *PLoS ONE* 7(3): e34222.

[8] Butler, David. J. and Graham C. Loomes (2007) "Imprecision as an Account of the Preference Reversal Phenomenon." *American Economic Review* 97(1): 277-297.

[9] Butler, David J. and Graham C. Loomes (2011) "Imprecision as an Account of Violations of Independence and Betweenness." *Journal of Economic Behavior & Organization* 80(3): 511-22.

[10] Chung, K. L. (1953) "Sur les Lois de Probabilité Unimodales." *Comptes Rendus Hebdomadaires des Séances de l'Académie de Sciences* 236(6): 583-584.

[11] Clark, James E. (2015) "Diet, exercise or diet with exercise: comparing the effectiveness of treatment options for weight-loss and changes in fitness for adults (18–65 years old) who are overfat, or obese; systematic review and meta-analysis." *Journal of Diabetes Metabolic Disorders* 14(1): Article 31.

[12] Cubitt, Robin P. , Daniel Navarro-Martinez and Chris Starmer (2015) "On preference imprecision." *Journal of Risk and Uncertainty* 50(1): 1-34.

[13] Danan, E., Gajdos, T., Hill, B. and J-M. Tallon (2016) "Robust Social Decisions." *American Economic Review* 106(9): 2407-2425.

[14] Dharmadhikari, Sudhakar, and Kumar Jogdeo (1976) "Multivariate Unimodality." *The Annals of Statistics* 4(3): 607-613.

[15] Dharmadhikari, Sudhakar, and Kumar Joag-Dev (1988) Unimodality, Convexity, and Applications. Elsevier.

[16] Fechner, G., (1860/1966). Elements of Psychophysics. New York: Holt, Rinehart and Winston.

[17] Georgescu-Roegen, Nicholas (1958) "Threshold in Choice and the Theory of Demand." *Econometrica* 26(1): 157-68.

[18] Ghosh, Pankaj (1974) "On Generalized Unimodality." *Communications in Statistics-Theory and Methods* 3(6): 567-580.

[19] Glenn, W. A. and H. A. David (1960) "Ties in Paired-Comparison Experiments Using a Modified Thurstone-Mosteller Model." *Biometrics* 16(1): 86-109.

[20] Greenberg, Marshall G. (1965) "A Modification of Thurstone's Law of Comparative Judgement to Accommodate a Judgement Category of "Equal" or "No Difference"." *Psychological Bulletin* 64(2): 108-112.

[21] Hodges, J. L. and E. L. Lehmann (1954) "Matching in Paired Comparisons." *Annals of Mathematical Statistics* 25(4): 787-791.

[22] Ibragimov, Il'dar Abdullovich (1956) "On the Composition of Unimodal Distributions." *Theory of Probability and its Applications* 1(2): 255-260.

[23] Kahneman, Daniel, Barbara L. Fredrickson, Charles A. Schreiber and Donald A. Redelmeier (1993) "When More Pain Is Preferred to Less: Adding a Better End." *Psychological Science* 4(6):401-405.

[24] Luce, Duncan R. (1956) "Semiorders and a Theory of Utility Discrimination." *Econometrica* 24(2): 178-191.

[25] Luce, Duncan R. (1959) Individual Choice Behavior: A Theoretical Analysis. New York: John Wiley and Sons.

[26] MacGillivray, H. L. (1981) "The Mean, Median, Mode Inequality and Skewness for a Class of Densities." *Australian Journal of Statistics* 23(2): 247-250.

[27] Mandelbrot, Benoit (1963) "New Methods in Statistical Economics." *Journal of Political Economy* 71(5): 421-440.

[28] Natenzon, Paulo (2016) "Random Choice and Learning." forthcoming in the *Journal of Political Economy.*

[29] Nadarajah, Saralees (2007) "Linear Combination of Gumbel Random Variables." *Stochastic Environmental Research and Risk Assessment* 21(3): 283-286.

[30] Permana, Yudistira (2018) "Why Do People Say They Are Unsure? An Experimental Investigation." Mimeo, University of York.

[31] Purkayastha, Sumitra (1998) "Simple proofs of two results on convolutions of unimodal distributions." *Statistics & Probability Letters* 39(2): 97-100.

[32] Quandt, Richard E. (1956) "A Probabilistic Theory of Consumer Behavior." *The Quarterly Journal of Economics*, 70(4): 507-536.

[33] Rao, P. V. and L. L. Kupper (1967) "Ties in Paired-Comparison Experiments: A Generalization of the Bradley-Terry Model." *Journal of the American Statistical Association*, 62(317):194-204.

[34] Ratcliff, R. (1978) "A theory of memory retrieval." *Psychological Review* 85: 59-108.

[35] Schünemann, Holger, Jan Brożek, Gordon Guyatt, and Andrew Oxman, editors (2013) "GRADE handbook for grading quality of evidence and strength of recommendations." Updated October 2013. The GRADE Working Group, 2013. Available from guidelinedevelopment.org/handbook.

[36] Thurstone, Louis L. (1927) "Psychophysical Analysis." *American Journal of Psychology* 38(3): 368-389.

[37] Thurstone, Louis L. (1927) "A Law of Comparative Judgement." *Psychological Review* 34: 273-286.

[38] Tyson, Christopher J. (2018) "Exponential Satisficing." Mimeo, Queen Mary University of London.

[39] Vogt, H. (1983) "Unimodality of Differences." *Metrika* 30(1): 165-170.

[40] Wintner, A. (1938) Asymptotic Distributions and Infinite Convolutions. Edwards Brothers.

[41] Yamazato, Makato (1978) "Unimodality of Infinitely Divisible Distribution Functions of Class L." *Annals of Probability* 6(4): 523-531.

[42] van Zwet, Willem R. (1979) "Mean, median, mode II." *Statistica Neerlandica*, 33(1): 1-5.

A Examples 2 and 3

In this Appendix, we follow the excellent treatment of Nadarajah [29]. When u_i and u_j are independent random variables, iterated expectations imply that

$$F_{u_i-u_j}(x) = \int_{\mathbb{R}} \Pr[u_i \leq z + x|z] f_j(z)\,dz = \int_{\mathbb{R}} F_i(z+x) f_j(z)\,dz. \tag{11}$$

In Examples 2 and 3, ε_i and ε_j are i.i.d. Gumbel with location $\nu_i = 0 = \nu_j$ and scale $c_i = c \geq c_j = 1$. So, u_i is Gumbel with location $\hat{u}_i$ and scale c while u_j is Gumbel with location $\hat{u}_j$ and scale 1. In other words,

$$F_i(w) = e^{-e^{-\frac{w-\hat{u}_i}{c}}} \text{ and } f_j(w) = e^{-e^{-(w-\hat{u}_j)}} e^{-(w-\hat{u}_j)}.$$

By replacing these formulas into (11) and defining $A(x) := e^{x-(\hat{u}_i-\hat{u}_j)}$, one obtains the cdf

$$F_{u_i-u_j}(x) = \int_{\mathbb{R}} e^{-e^{-\frac{z+x-\hat{u}_i}{c}}} e^{-e^{-(z-\hat{u}_j)}} e^{-(z-\hat{u}_j)} dz = cA(x) \int_{\mathbb{R}_+} y^{c-1} e^{-(A(x)y^c+y)} dy \qquad (12)$$

where the last equality follows from the change of variables $y = e^{-\frac{z+x-\hat{u}_i}{c}}$.

By differentiating (12) with respect to x, one obtains the density

$$f_{u_i-u_j}(x) = cA(x) \int_{\mathbb{R}_+} y^{c-1} e^{-(A(x)y^c+y)} dy - cA(x)^2 \int_{\mathbb{R}_+} y^{2c-1} e^{-(A(x)y^c+y)} dy. \qquad (13)$$

A.1 Example 2

In the special case where $c = 1$, the expressions in (12) and (13) simplify to

$$F_{u_i-u_j}(x) = A(x) \int_{\mathbb{R}_+} e^{-y(A(x)+1)} dy = \frac{A(x)}{A(x)+1} \qquad (14)$$

$$f_{u_i-u_j}(x) = A(x) \int_{\mathbb{R}_+} e^{-y(A(x)+1)} dy - A(x)^2 \int_{\mathbb{R}_+} y e^{-y(A(x)+1)} dy = \frac{A(x)}{(A(x)+1)^2} \qquad (15)$$

(The last equality in (15) follows from integration by parts.) These formulas correspond to the cdf and density of a logistic distribution with location $\hat{u}_i - \hat{u}_j$ and scale 1.

Given (14), the probability that i "beats" j is then given by

$$\Pr[u_i > u_j + \tau] = 1 - F_{u_i-u_j}(\tau) = \frac{1}{A(\tau)+1} = \frac{e^{\hat{u}_i}}{e^{\hat{u}_j+\tau} + e^{\hat{u}_i}}.$$

Using this formula, the analysis then proceeds as in the main text.

A.2 Example 3

When $c > 1$, simple closed form expressions are lacking and the analysis is much more involved. Nonetheless, we can make the following simple observations:

(i) Since u_i and u_i are independent Gumbel, their mean value difference is

$$\mathbb{E}[u_i - u_j] = \mathbb{E}[u_i] - \mathbb{E}[u_j] = [\hat{u}_i + c\gamma] - [\hat{u}_j + \gamma] = (\hat{u}_i - \hat{u}_j) + (c-1)\gamma$$

(where, as in the text, γ denotes the Euler-Mascheroni constant).

(ii) For c sufficiently close to 1, the median of the value difference $med[u_i - u_j]$ is approximated by the difference of the median values so that

$$med[u_i - u_j] \approx med[u_i] - med[u_j] = [\hat{u}_i - c \ln\ln 2] - [\hat{u}_j - \ln\ln 2] = (\hat{u}_i - \hat{u}_j) + (c-1)|\ln\ln 2|.$$

From (i) and (ii), it follows that i is both mean-better *and* median-worse than j when

$$0.36651 \approx |\ln\ln 2| < \frac{\hat{u}_j - \hat{u}_i}{c-1} < \gamma \approx 0.57722.$$

B Proof of Proposition 1

First, define $M := 2\max\{\hat{\mu}, \hat{m}\}$. For $\varepsilon > 0$, let $T_{(c,\varepsilon)}$ denote the symmetric triangular distribution centred at c with support on the interval $[c-\varepsilon, c+\varepsilon]$; and let $t_{(c,\varepsilon)}$ denote its density. Finally, let $u_i - u_j$ denote the even mixture between $T_{(-\hat{\tau},\varepsilon)}$ and $T_{(M+\hat{\tau},\varepsilon)}$; and let $f_{u_i-u_j} := \frac{t_{(-\hat{\tau},\varepsilon)}+t_{(M+\hat{\tau},\varepsilon)}}{2}$ denote the density of this mixture distribution.

By symmetry, $\mathbb{E}[u_i - u_j] = med[u_i - u_j] = \frac{M}{2} \geq \hat{\mu}, \hat{m}$. What is more,

$$f_{u_i-u_j}(-\hat{\tau}) = \frac{t_{(-\hat{\tau},\varepsilon)}(-\hat{\tau})}{2} = \frac{1}{\varepsilon} \geq 2\delta \quad \text{and} \quad f_{u_i-u_j}(\hat{\tau}) = 0$$

by choosing $0 < \varepsilon \leq \min\{M, 2\hat{\tau}, \frac{1}{2\delta}\}$. Differentiating (2) it then follows that

$$\frac{\partial p\,(i,\hat{\tau})}{\partial \tau} = \frac{1}{2}\left[f_{u_i-u_j}\,(-\hat{\tau}) - f_{u_i-u_j}\,(\hat{\tau})\right] = \frac{1}{2\varepsilon} \geq \delta.$$

With the specified choices of M and ε, $f_{u_i-u_j}$ satisfies all of the desired requirements.

C Unimodality: classical results

If u_i and u_j are independent with densities denoted by f_i and f_j, then (3) simplifies to

$$f_{u_i-u_j}\,(x) = \int_{\mathbb{R}} f_i\,(z+x)\,f_j\,(z)\,dz = \int_{\mathbb{R}} f_i\,(z)\,f_j\,(z-x)\,dz \quad \text{for all } x \in \mathbb{R}. \tag{16}$$

This formula may be viewed as a convolution of two densities. To see this, recall that the *convolution* $h * g$ of two functions $g, h : \mathbb{R} \to \mathbb{R}$ is defined by

$$[h*g](x) := \int_{\mathbb{R}} h\,(z)\,g\,(x-z)\,dz \quad \text{for all } x \in \mathbb{R}.$$

It is immediate from the definition that convolution has some nice algebraic properties—including that it is (i) commutative, (ii) associative and (iii) commutes with translation.[30] Since it will be quite useful in the sequel, we denote the *reflection* of a function $f : \mathbb{R} \to \mathbb{R}$ through zero by $\bar{f}$. In other words, $\bar{f}\,(y) := f\,(-y)$ for all $y \in \mathbb{R}$. (Notice that the function f is symmetric around zero if and only if $f = \bar{f}$.[31])

Using this notation, equation (16) simplifies to

$$f_{u_i-u_j}\,(x) = [f_i * \bar{f}_j]\,(x)\,. \tag{17}$$

Having expressed $f_{u_i-u_j}$ as a convolution of densities, we are in position to exploit some classical results about the preservation of unimodality under convolution.

The first result, due to Ibragimov [22], states that unimodality is preserved under convolution provided that *one* of the densities is log-concave. Since reflection preserves unimodality and log-concavity, formula (17) allows us to re-state his result as follows:

[30]Formally: (i) $h*g = g*h$, (ii) $h*(g*f) = (h*g)*f$ and (iii) $(\tau_\nu h)*g = \tau_\nu(h*g)$. Recall that the translation of a function $f : \mathbb{R} \to \mathbb{R}$ by $c \in \mathbb{R}$ is given by $[\tau_c f](x) := f(x+c)$ for all $x \in \mathbb{R}$.

[31]Symmetry of f around $c \in \mathbb{R}$ amounts to the symmetry of the translation $\tau_{-c}f$ around zero.

Theorem 1. ***(Ibragimov)*** *Suppose that u_i and u_j are independently distributed random variables.* ***(I)*** *If one is unimodal and the other is log-concave, then $f_{u_i-u_j}$ is unimodal.* ***(II)*** *If the unimodal (but not necessarily log-concave) random variable is* also *strictly unimodal, then $f_{u_i-u_j}$ is strictly unimodal.*

Technically, Ibragimov's result only implies (I). To show (II), we adapt the argument that Dharmadhikari and Joag-Dev [15] use in the proof of their Theorem 1.10(a):

Proof of (II): Without loss of generality, suppose that f_i is strictly unimodal around $\bar{\nu}_i = 0$ and f_j is log-concave around $\bar{\nu}_j = 0$. (The argument is symmetric when f_j is strictly unimodal and f_i is log-concave; and, when the central modes of u_i and u_j are non-zero, it can be applied directly to the "de-moded" random variables.[32]) For simplicity, we also suppose that f_i is differentiable and $f_j(z) > 0$ for all $z \in \mathbb{R}$. (One can remove these restrictions using Dharmadhikari and Joag-Dev's arguments from the proof of their Theorem 1.10(a).) Under these restrictions, formula (16) implies

$$f'_{u_i-u_j}(x) = \int_{\mathbb{R}} f'_i(z) f_j(z-x)dz$$

for all $x \in \mathbb{R}$. The crux of the proof is then to establish that

$$\bar{f}_j(y) f'_{u_i-u_j}(w) > \bar{f}_j(w) f'_{u_i-u_j}(y) \tag{18}$$

for all $y > w$. To see that the desired result follows from (18), first note that $f_{u_i-u_j}$ is unimodal around $\bar{\nu} = 0$ by Theorem 1(I). So, the stated restrictions on f_i and f_j imply that $f'_{u_i-u_j}(0) = 0$. Substituting $(y, w) = (0, x)$ and $(y, w) = (x, 0)$ into (18) gives

$$f'_{u_i-u_j}(x) \begin{cases} < 0 & \text{for } x \in (0, +\infty) \\ = 0 & \text{for } x = 0\text{, and} \\ > 0 & \text{for } x \in (-\infty, 0). \end{cases}$$

In other words, $f_{u_i-u_j}$ is strictly unimodal around zero, which is the desired result.

To establish inequality (18), fix some $w, y \in \mathbb{R}$ such that $y > w$. Then, since f_j is log-concave and f_i is strictly unimodal, the following inequalities hold for any $z > 0$:

$$f_j(y) f_j(z-w) \leq f_j(w) f_j(z-y) \quad \text{and} \quad f'_i(z) < 0. \tag{19}$$

By combining the two inequalities in (19) and integrating over $\mathbb{R}_+$, one obtains

$$f_j(y) \int_{\mathbb{R}_+} f'_i(z) f_j(z-w)dz > f_j(w) \int_{\mathbb{R}_+} f'_i(z) f_j(z-y)dz. \tag{20}$$

[32]In particular, $(\tau_{-\bar{\nu}_i} f_i) * \overline{(\tau_{-\bar{\nu}_j} f_j)} = (\tau_{-\bar{\nu}_i} f_i) * (\tau_{\bar{\nu}_j} \bar{f}_j) = (\tau_{-\bar{\nu}_i} \cdot \tau_{\bar{\nu}_j})[f_i * \bar{f}_j] = \tau_{\bar{\nu}_j - \bar{\nu}_i}[f_i * \bar{f}_j]$.

Since the inequalities in (19) are both reversed when $z < 0$, the same argument gives

$$f_j(y) \int_{\mathbb{R}_-} f_i'(z) f_j(z-w) dz > f_j(w) \int_{\mathbb{R}_-} f_i'(z) f_j(z-y) dz. \tag{21}$$

By combining (20) and (21), one then obtains the desired inequality (18). ■

Before Ibragimov, Wintner [40] had already shown that unimodality is preserved under convolution when *both* of the distributions are symmetric.[33] Since reflection preserves symmetry, formula (17) allows us to re-state his result as follows.

Theorem 2. ***(Wintner)*** *Suppose that u_i and u_j are independently distributed random variables that are symmetric around $\bar{\nu}_i$ and $\bar{\nu}_j$, respectively.* ***(I)*** *If u_i and u_j are unimodal, then $f_{u_i-u_j}$ is unimodal. What is more, $f_{u_i-u_j}$ is symmetric around $\bar{\nu}_i - \bar{\nu}_j$.* ***(II)*** *If one of the two random variables is <u>also</u> strictly unimodal, then $f_{u_i-u_j}$ is strictly unimodal.*

Wintner's result only implies the first sentence in (I). For the second sentence, suppose $\bar{\nu}_i = 0 = \bar{\nu}_j$. (By the argument in Theorem 1(II), this is without loss.) Then,

$$f_{u_i-u_j}(x) = [f_i * \bar{f}_j](x) = [\bar{f}_i * f_j](x) = [f_j * \bar{f}_i](x) = f_{u_j-u_i}(x)$$

where the second equality holds by symmetry and the third by commutativity. Since $f_{u_j-u_i}(x) = f_{u_i-u_j}(-x)$ for all $x \in \mathbb{R}$, $f_{u_i-u_j}$ is symmetric around zero.

For (II), we use Theorem 1.5(b) of Dharmadhikari and Joag-Dev [15], which shows that: *the set of random variables that are unimodal and symmetric around zero coincides with the convex hull of uniform random variables that are symmetric around zero.*

Proof of (II): Without loss of generality, suppose that u_i is strictly unimodal and $\bar{\nu}_i = 0 = \bar{\nu}_j$. By the cited result of Dharmadhikari and Joag-Dev, it then suffices to establish that $f_{u_i-u_j}$ is strictly unimodal when u_j is uniform on $[-a, a]$ for some $a \in \mathbb{R}_+$. To show this, first note that $f_i(z) > 0$ for all $z \in \mathbb{R}$ (by strict unimodality of u_i); and $f_j(z) = \frac{1}{2a}$ for all $z \in [-a, a]$ (by uniformity of u_j). From (16), it then follows that

$$f_{u_i-u_j}(x) = \int_{\mathbb{R}} f_i(z+x) f_j(z) dz = \frac{1}{2a} \int_{-a}^{a} f_i(z+x) dz = \frac{F_i(x+a) - F_i(x-a)}{2a} \tag{22}$$

for all $x \in \mathbb{R}$. Differentiating (22) and using the symmetry of f_i around zero gives

$$f'_{u_i-u_j}(x) = \frac{f_i(x+a) - f_i(x-a)}{2a} = \frac{f_i(|x+a|) - f_i(|x-a|)}{2a}. \tag{23}$$

Since f_i is strictly unimodal, it is decreasing on $(0, +\infty)$. Using (23), it follows that

$$f'_{u_i-u_j}(x) \begin{cases} < 0 & \text{for } x \in (0, +\infty) \\ = 0 & \text{for } x = 0, \text{ and} \\ > 0 & \text{for } x \in (-\infty, 0). \end{cases}$$

[33] For a concise and recent treatment of this result, see Purkayastha [31] (Theorem 2.1).

In other words, $f_{u_i - u_j}$ is strictly unimodal around zero, which is the desired result. ■

A third important result about unimodality is Hodges and Lehmann's [21] observation that the convolution of a unimodal density f with its reflection $\bar{f}$ is unimodal (see also Purkayastha [31] (Theorem 2.2), Dharmadhikari and Joag-Dev [15] (Theorem 1.8) or Vogt [39]). (In statistics, the convolution $f * \bar{f}$ is known as the *symmetrization* of f.) Using formula (17), it is possible to translate their result into our framework as follows:

Theorem 3. ***(Hodges and Lehmann)*** *Suppose that u_i and u_j are unimodal i.i.d. random variables. Then, $f_{u_i - u_j}$ is strictly unimodal and symmetric around zero.*

Hodges and Lehmann's result is equivalent to the unimodality of $f_{u_i - u_j}$ around zero. In turn, the symmetry of $f_{u_i - u_j}$ follows from the same kind of argument used to establish the second sentence of Theorem 2(I) above. Finally, the strict unimodality of $f_{u_i - u_j}$ follows from an application of the Cauchy-Schwarz inequality. In particular:

Proof: Since $f_{u_i - u_j}$ is unimodal and symmetric at zero, it suffices to show that $f_{u_i - u_j}(0) > f_{u_i - u_j}(x)$ for all $x > 0$. Towards a contradiction, suppose that $f_{u_i - u_j}(x) = f_{u_i - u_j}(0)$ for some $x > 0$. Then, from equation (16), it follows that

$$\int_{\mathbb{R}} f(z + x) f(z)\, dz = f_{u_i - u_j}(x) = f_{u_i - u_j}(0) = \int_{\mathbb{R}} f^2(z)\, dz$$

where f is the density of u_i. By manipulating the right-hand side, one obtains

$$\int_{\mathbb{R}} f(z + x) f(z)\, dz = \left(\sqrt{\int_{\mathbb{R}} f^2(z)\, dz}\right)^2 = \sqrt{\int_{\mathbb{R}} f^2(z + x)\, dz} \cdot \sqrt{\int_{\mathbb{R}} f^2(z)\, dz}. \qquad (24)$$

By the Cauchy-Schwarz inequality, equation (24) implies that $f(z) = f(z + x)$ for almost all $z \in \mathbb{R}$.[34] By the argument given at the end of the proof of Proposition 10, it then follows that $\int_{\mathbb{R}} f(z)dz \neq 1$. But, this contradicts the fact that f is a density. ■

D Three alternatives

To derive an explicit formula for (9), one must compute R_i^τ and R_{ij}^τ for $i, j \in \{1, 2, 3\}$. Where $f(x, y, z)$ denotes the joint density of (u_1, u_2, u_3), it is straightforward to see that

$$R_1^\tau = \int_{\mathbb{R}} \int_{-\infty}^{x-\tau} \int_{-\infty}^{x-\tau} f(x, y, z)\, dydzdx \qquad (25)$$

$$R_{12}^\tau = \int_{\mathbb{R}} \int_{x-\tau}^{x} \int_{-\infty}^{x-\tau} f(x, y, z)\, dzdydx + \int_{\mathbb{R}} \int_{y-\tau}^{y} \int_{-\infty}^{y-\tau} f(x, y, z)\, dzdxdy \qquad (26)$$

[34] On its own, the Cauchy-Schwarz inequality only implies that there exists some $\lambda \in \mathbb{R}$ such that $f(z) = \lambda f(z + x)$ for almost all $z \in \mathbb{R}$. Since $\int_{\mathbb{R}} f(z)dz = \int_{\mathbb{R}} f(z + x)dz$ however, it follows that $\lambda = 1$.

The expressions for R_2^τ, R_3^τ, R_{13}^τ and R_{23}^τ are symmetric. To elaborate, observe that $\{u_1\}$ is the top-set if and only if the value realizations are such that $u_2, u_3 < u_1 - \tau$, which gives (25). In turn, (26) follows from the observation that $\{u_1, u_2\}$ is the top-set if and only if: (i) $u_3 < u_1 - \tau \leq u_2 \leq u_1$; or, similarly, (ii) $u_3 < u_2 - \tau \leq u_1 \leq u_2$.

To derive an explicit formula for (10), one must compute the threshold probabilities $R_{i\to ij}^\tau$ and $R_{ij\to ijk}^\tau$ for $i, j, k \in \{1, 2, 3\}$. By the same kind of reasoning as above,

$$R_{2\to 12}^\tau = \int_{\mathbb{R}} \int_{-\infty}^{y-\tau} f\left(y - \tau, y, z\right) dzdy \tag{27}$$

$$R_{23\to 123}^\tau = \int_{\mathbb{R}} \int_{y-\tau}^{y} f\left(y - \tau, y, z\right) dzdy + \int_{\mathbb{R}} \int_{z-\tau}^{z} f\left(z - \tau, y, z\right) dydz \tag{28}$$

The other threshold probabilities are symmetric. For (27), note that the boundary between top-sets $\{u_2\}$ and $\{u_1, u_2\}$ is defined by the requirement that $u_3 < u_1 = u_2 - \tau$. Similarly, for (26), note that the boundary between top-sets $\{u_2, u_3\}$ and $\{u_1, u_2, u_3\}$ requires: (i) $u_1 = u_2 - \tau \leq u_3 \leq u_2$; or (ii) $u_1 = u_3 - \tau \leq u_2 \leq u_3$.

To obtain an explicit formula for $p(i, \tau)$, one can replace (25), (26) and their analogs into equation (9). One can then check that the result obtained by differentiating $p(i, \tau)$ with respect to τ coincides exactly with the formula given by replacing (27), (28) and their analogs into equation (10). We leave these calculations to the reader.