Theoretical Economics 15 (2020), 989–1022 1555-7561/20200989

Twisting the truth: Foundations of wishful thinking

Matthew Kovach
Department of Economics, Virginia Tech

Considerable evidence shows that people have optimistic beliefs about future outcomes. I present an axiomatic model of wishful thinking (WT), in which an endowed alternative, or status quo, influences the agent's beliefs over states and thus induces such optimism. I introduce a behavioral axiom formalizing WT and derive a representation in which the agent overweights states in which the endowment provides a higher payoff. WT is a novel channel through which an endowment may influence choice behavior and provides a coherent explanation for a variety of observed behavior, including choice reversals among nonstatus quo alternatives when the status quo changes. WT leads to inefficient risk sharing in an exchange economy and has unique implications for the gap between willingness to accept and willingness to pay for endowed goods.

Keywords. Wishful thinking, status quo bias, reference dependence, belief distortions, optimism.

JEL classification. D01, D11, D80, D81, D83.

1. Introduction

Every day a multitude of decisions are influenced by the presence of a reference point. This influence on choice behavior is abundantly evidenced by experimental and empirical studies in economics and psychology. One type of reference point has received special attention in the context of economic decision making: the endowment (also referred to as the status quo or default option). This attention to the endowment is partially due to how readily it can be observed in many instances. But more significantly, it deserves this attention because most real-life decisions feature an endowment or status quo. A few real-life examples include deciding how to adjust current investments, such as a 401(k) (see Samuelson and Zeckhauser 1988), changing jobs, and buying an insurance plan.

Matthew Kovach: mkovach@vt.edu
I would like to especially thank Federico Echenique and Pietro Ortoleva for their support and guidance. Two excellent referees provided many thoughtful comments that substantially improved this paper. I also thank Kota Saito, Leeat Yariv, Roland Bénabou, Simone Cerreia-Vioglio, Matthew Chao, Jonathan Chapman, Liam Clegg, William Fuchs, Philip Sadowski, Tomasz Strzalecki, Gerelt Tserenjigmid, and seminar participants at ITAM, University of Haifa, Midwest Economic Theory Conference 2014, SWET 2015, and RUD 2015. This paper is based on a chapter from my doctoral dissertation at Caltech and was previously circulated as "Twisting the Truth: A Theory of Cognitive Dissonance." All errors are my own.

© 2020 The Author. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at https://econtheory.org. https://doi.org/10.3982/TE2744

Existing models of status quo bias have focused on the decision maker choosing the status quo too often, revealing excessive persistence.[1] A common feature of these models is that the status quo creates a mental constraint, and thus restricts the alternatives the agent considers choosable. Further, for choice problems in which the agent abandons the status quo at every submenu, the status quo is "irrelevant" and, therefore, his choice must be identical to choice without a status quo. That is, the status quo exerts a pull on the agent without otherwise distorting preferences.

However, many of the existing models are inconsistent with two striking behavioral patterns. First, a change in the status quo often induces choice reversals among non-status quo alternatives.[2] For example, consider the choice between a sure payoff, S, a low-risk gamble, L, and a riskier gamble, R. With no status quo, the agent ranks alternatives $S \succ L \succ R$ and chooses S. But when L is the status quo, the agent ranks alternatives $R \succ_L L \succ_L S$ and chooses R. This choice pattern is a form of *Generalized Status Quo Bias* and is consistent with behavior observed in Dean et al. (2017) and Masatlioglu and Uler (2013).[3] Second, beliefs about future outcomes tend to positively align with the agent's "current situation." For example, Mayraz (2011b) randomly endowed subjects with a role as a farmer or baker and found that the subjects' beliefs were systematically skewed in favor of the subjects' endowed role.

To explain these behavioral patterns, I propose a different channel through which an endowment affects behavior in environments with uncertainty: *wishful thinking* (see Section 1.1 for evidence). Within a standard environment for decision making under uncertainty (see Section 2), I consider a system of preference relations for the agent: one for each endowment and one representing "no endowment." Given endowment f, a wishful thinker shifts beliefs so that states in which f yields relatively higher payoffs are more likely. Now consider any act which is "aligned" with f: it yields relatively higher payoffs in the same states which f does. Since wishful thinking operates through beliefs, any such act would also be viewed more favorably when endowed with f than it would be otherwise. In Section 3 I introduce a behavioral axiom which imposes a preference for acts aligned with the endowment and show that this axiom, along with standard conditions, characterizes a general model of wishful thinking. Wishful thinkers are subjective expected utility maximizers with endowment-dependent beliefs. Conditional on endowment f, the belief in state s is $\mu_f(s) = \mu(s)\delta_f(u(f(s)))$, where μ is the reference-free belief, u is a utility index, and δ_f is an increasing "distortion" function. This amounts to a "twisting" of the indifference curves in utility space; hence preferences are globally dependent on the endowment. This twisting of indifference curves generates the generalized status quo bias choice pattern.

[1] Kahneman and Tversky (1979) were among the first to point out the importance of the status quo in determining choice behavior. For recent examples in decision theory, see Masatlioglu and Ok (2005, 2014) and Ortoleva (2010).

[2] Dean et al. (2017) and Eliaz and Spiegler (2011) can accommodate some reversals of this type through limited attention. In general, limited attention is conceptually and behaviorally distinct from wishful thinking.

[3] This choice pattern also features a "decrease" in risk aversion, as was found in Sprenger (2015).

Since δ_f is endowment-dependent, behavior may vary substantially across different endowments. In order to generate sharper predictions and facilitate application, I characterize two special cases that impose substantial structure on distortions. In the first case, the Consequential distortion, the relative likelihood between any two states only depends on the payoff of f in those states; δ_f is determined by some increasing function v and a normalizing constant. The Consequential distortion is characterized by the addition of an independence property, Independence of irrelevant payoffs: whenever two endowments provide the same state-wise payoffs on some event, the agent's ranking of acts that vary only on that event are consistent across endowments. In the second case, the Best-case binary distortion, beliefs are a convex combination of a reference-free belief and a belief that maximizes the value of the status quo. Hence there are *good* states, determined by f, which are given additional weight relative to *bad* states, while other ratios are unchanged. While this is "more restrictive" than the Consequential distortion, as it allows fewer distortions to relative probabilities, it is "less restrictive" as it allows the notion of good states and their additional weight to depend on f. This representation also requires a single additional axiom, Best-case dominance: if the reference-free preference and a "maximally wishful" preference agree on their ranking of h and g, and both are aligned with f, then this ranking also holds for preferences conditional on f. The special cases discussed above are in Section 4 and Section 5, respectively. These two special cases are essentially disjoint: imposing both axioms together eliminates wishful thinking and results in standard expected utility behavior, as I demonstrate in Section 6.

To facilitate comparative statics in applications, I propose a comparative notion of wishful thinking and formalize its behavioral content for the Consequential distortion and the Best-case dominance in Section 7. Say that agent one is *more wishful* than agent two if whenever agent two is unwilling to abandon f for some sure payoff, then so is agent one. This result supports the interpretation of model parameters and guides the specification of parametric families of distortions which are ordered by their degree of wishful thinking. I then provide two applications of wishful thinking in Section 8. The first application considers the implications of wishful thinking for asset prices and risk sharing in a simple exchange economy. This application highlights a unique consequence of inequality and shows that typically equilibria with wishful thinking involve less than full risk sharing; the more wishful thinking an agent exhibits, the more risk he holds. The second application considers the well-known gap between willingness to accept and willingness to pay. Wishful thinking preferences predict that (i) a gap exists only for uncertain alternatives, (ii) the gap increases for mean preserving spreads and (iii) the gap is increasing in the degree of wishful thinking.

I close by discussing related literature in Section 9, including formal comparisons to the most closely related papers. Of particularly close relation is Mayraz (2011a), which characterizes a similar representation to mine. In his paper, the belief distortion, $\delta_f(u(f(s)))$, takes a logistic form so that $\frac{\delta_f(u(f(s)))}{\delta_f(u(f(\tilde{s})))} = e^{\lambda[u(f(s))-u(f(\tilde{s}))]}$. His characterization uses different primitives and axioms, and a careful discussion appears in Section 9.2. I also compare wishful thinking to models of status quo bias (see Masatlioglu and Ok 2005 or its generalization Masatlioglu and Ok 2014) and demonstrate that wishful thinking is behaviorally distinct from the model of Masatlioglu and Ok (2014) (MO).

In MO, the canonical model of status quo bias, the agent behaves as follows: given a choice set A, when f is a status quo the agent maximizes $\succsim$ over $A \cap \mathcal{Q}(f)$. Hence the status quo determines a constraint set over which the agent maximizes a reference-free preference. I show that the model of wishful thinking I derive and the model in MO are essentially disjoint. Imposing either one of the axioms from MO eliminates all effects of the endowment. This is because a wishful thinker twists his preferences, while in MO the agent always maximizes the same preference over status quo-dependent choice sets. This distinction is why wishful thinking accommodates the generalized status quo bias example discussed earlier, while MO does not.

1.1 *Wishful thinking: Experimental and empirical evidence*

Optimism comes in two forms: optimism about one's own performance or ability, or optimism about future events over which a decision maker has no control. The first form, overconfidence, has received much attention in the economics literature while the second, wishful thinking, has received relatively little. Nevertheless, a variety of psychology and neuroscience experiments have found that subjects have optimistic and stakes-dependent beliefs. In one prominent study, Weinstein (1980) found that subjects have unrealistically optimistic views about their future outcomes, including job prospects, earnings, and health outcomes. Sharot et al. (2007) and Sharot (2011) suggest that this optimism bias is a product of normal brain function and can also be observed in other animals, suggesting strong evolutionary origins. In addition to such a neurological mechanism, other psychological phenomena that may contribute to wishful thinking include motivated reasoning (Kunda 1987) and illusion of control (Langer 1975 and Budescu and Bruderman 1995). The relative importance of these forces in generating wishful thinking is interesting, but further speculation as to the cause of wishful thinking is beyond the scope of this paper. The rest of this section will introduce some of the experimental and empirical evidence of wishful thinking in more economic contexts.

In the lab: Mijović-Prelec and Prelec (2010) asked subjects to make incentivized predictions about binary events, both before and after being randomly assigned payoffs that depend on the outcomes. They found that subjects adjusted their predictions to increase the probability of the higher payoff event, which is inconsistent with rational expectations but in line with wishful thinking. Similarly, Mayraz (2011b) found that assigning subjects to opposite sides of the market caused subjects' beliefs to diverge, each overestimating his or her future profit (see also Babcock et al. 1995). Both of these experiments illustrate a tendency for "beliefs to follow payoffs."

Out of the lab: DiTella et al. (2007) documented the "pro-Market" beliefs of squatters outside Buenos Aires after some (exogenously) received property rights. There was a large reported difference in the beliefs of those with and without property rights, despite the fact that they lead nearly identical lives and lived in close proximity to each other. The findings in DiTella et al. (2007) indicate that more general types of beliefs are "malleable" and adjust to support one's current situation. Cohen (2009) found that employees severely overinvested in employer stock and consequently suffered a near 20% reduction in retirement income. Since employees tended to increase investment in employer stock after a spin-off, rather than just hold at their current levels, this effect is

likely not driven by a mere reluctance to abandon a status quo. These (costly) increases in investment suggest that the post spin-off beliefs were overly optimistic, consistent with wishful thinking.

2. Setup and foundations

I adopt a standard setup for studying the effect of information on preferences. There is a finite set S of states of the world, with $|S| \geq 2$.[4] Events are denoted $A, B, C \in \Sigma = 2^S \backslash \{\varnothing\}$ and X denotes the set of consequences. Let $\mathcal{F}$ denote the set of all acts, which are functions $f : S \to X$. Following a standard abuse of notation, let $x \in \mathcal{F}$ denote the constant act that returns $x \in X$ in every state. For any event A and acts $f, g \in \mathcal{F}$, let fAg denote the composite act *f on A, g otherwise*: $fAg(s) = f(s)$ if $s \in A$ and $g(s)$ if $s \in A^c$.

I assume that X is a convex subset of a vector space (see Maccheroni et al. 2006). For example, X could be the set of monetary prizes (e.g., $X = \mathbb{R}_+$) or X could be the set of lotteries over some set Y (which corresponds to the classic setup of Anscombe and Aumann 1963). This assumption on X allows mixtures to be defined in the usual way: for every $f, g \in \mathcal{F}$, and $\alpha \in [0, 1]$, the mixed act $\alpha f + (1-\alpha)g \in \mathcal{F}$ is that act returning the prize $\alpha f(s) + (1-\alpha)g(s) \in X$ for every $s \in S$.

I take the collection of preference relations $\{\succsim, \succsim_f\}_{f \in \mathcal{F}}$ over the set of acts as a primitive. Here, $\succsim$ represents the agent's neutral preferences,[5] while $\succsim_f$ is interpreted as preference given f, where f is an endowment or status quo. This setting matches that of Masatlioglu and Ok (2005, 2014), except that I take preference as a primitive rather than a choice correspondence.

2.1 *Axioms*

To isolate the effects of endowment-induced wishful thinking on behavior, I assume that once the impact of the endowment is considered the agent is otherwise standard. Consequently, conditional on an endowment, the agent is assumed to be a subjective expected utility (SEU) maximizer. The first axiom, Consistent expected utility, is a collection of well-known postulates (found in Appendix A) and imposes precisely this assumption.

Axiom 1 (Consistent expected utility). *$\succsim$ and $\succsim_f$ are subjective expected utility preferences for all $f \in \mathcal{F}$.*

Such an approach is not without criticism, since it is plausible that an agent susceptible to wishful thinking may also exhibit a multitude of other biases. The goal of this paper is to behaviorally understand wishful thinking, not wishful thinking compounded

[4]The assumption of finite S is merely for convenience. All results are unchanged if I assume an infinite state space and restrict attention to nonnull events. What is crucial is the existence of at least two nonnull events.

[5]In the choice literature on status quo bias, the symbol $\diamond$ is often used to denote neutral choice, e.g., $C(M, \diamond)$ is the choice from menu M without a status quo. $\succsim$ plays an identical role in this paper.

with other biases. With this in mind, the assumption of SEU preferences is the natural starting point, and I leave it to future work to study the effects of compounding biases.

Wishful thinking refers to the tendency for beliefs about future outcomes to align with an agent's current situation. Behaviorally, this is equivalent to a preference for acts that are "in alignment with the endowment." Formally, xAz is more aligned with f than xBz when A is a clearly better event under f: $f(s) \succsim f(\tilde{s})$ for all $s \in A$ and $\tilde{s} \in B$. That is, the agent prefers acts that yield better payoffs in the same states in which the endowment also yields better payoffs. For example, an agent might prefer a car with an electric motor, e, over one with a gasoline motor, g, while a similar agent who recently inherited shares in an oil company, o, might prefer g.[6] This is because wishful thinking leads the agent with oil investments to believe that future environmental regulations are less likely. Hence wishful thinking allows for the following preference reversal: $e \succ g$ but $g \succ_o e$. I now introduce the main behavioral axiom which formalizes this idea.

Axiom 2 (Wishful Thinking). *For all $f \in \mathcal{F}$, any $A, B \subset S$ such that for every $s \in A$ and $\tilde{s} \in B$, $f(s) \succsim f(\tilde{s})$, and all $x, y, z \in X$ such that $x, y \succsim z$,*

$$xAz \succsim yBz \quad \Longrightarrow \quad xAz \succsim_f yBz.$$

Wishful Thinking restricts attention to binary acts, in which case the notion of "alignment" is straightforward.

3. Representing wishful thinking

Definition 1 (Wishful Thinking Representation). An agent admits a *Wishful Thinking Representation* if there exists a utility function $u : X \to \mathbb{R}$, a belief $\mu \in \Delta(S)$, and for each f, an increasing distortion function $\delta_f : u(X) \to \mathbb{R}_+$ such that:

(i) $\succsim$ is represented by $V(g) = \sum_{s\in S} u(g(s))\mu(s)$, and

(ii) $\succsim_f$ is represented by $V_f(g) = \sum_{s\in S} u(g(s))\mu_f(s)$, where

$$\mu_f(s) = \delta_f\big(u\big(f(s)\big)\big)\mu(s).$$

In this case, say that $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ is a collection of *Wishful Thinking Preferences*.

This is the most general model of wishful thinking, where the distortion of particular states may vary considerably across endowments even if they have similar payoffs. Note that this embeds the standard model as a special case, where $\delta_f(a) = 1$ for all $a \in u(X)$ and $f \in \mathcal{F}$.

Theorem 1. *The following are equivalent:*

(i) $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ satisfy Consistent expected utility and Wishful Thinking.

(ii) The agent admits a Wishful Thinking Representation.

[6]Indeed, there is even a mild hedging motive in favor of e when invested in o, and thus wishful thinking is driving *anti-hedging* behavior. This is exactly in line with the findings in Cohen (2009).

Corollary 1. *If (u, μ, δ_f) and (u', μ', δ'_f) both represent the collection $\{\succsim, \succsim_f\}_{f \in \mathcal{F}}$, then $u' = \alpha u + \beta$ for $\alpha > 0$, $\beta \in \mathbb{R}$, $\mu' = \mu$ and $\delta'_f(u'(x)) = \delta_f(u(x))$ for all $x \in f(S)$.*

It is standard to show that the utility index u is unique up to a positive affine transformation and that μ and μ_f are unique. The uniqueness of δ_f follows from the uniqueness of μ_f and decomposition into the product $\delta_f(u(f(s)))\mu(s)$.

3.1 *Properties of wishful thinking*

Consider some endowment f. Then the relative likelihood of state s to state $\tilde{s}$ is given by

$$\frac{\mu_f(s)}{\mu_f(\tilde{s})} = \frac{\delta_f\big(u\big(f(s)\big)\big)}{\delta_f\big(u\big(f(\tilde{s})\big)\big)} \times \frac{\mu(s)}{\mu(\tilde{s})}.$$

Since δ_f is increasing, if $f(s) \succsim f(\tilde{s})$ then an agent endowed with f believes s to be *relatively* more likely than $\tilde{s}$ than without an endowment. Notice that if $f(s) \sim f(\tilde{s})$, then $\delta_f(u(f(s))) = \delta_f(u(f(\tilde{s})))$, and the relative likelihood between states that provide identical payoffs under f is undistorted. The intuition here is that however the agent feels about s, he should have precisely the same feelings about $\tilde{s}$ because they are equally good according to his endowment. Hence s and $\tilde{s}$ pull on the agent's beliefs in the same way.

Example 1. Suppose $X = (w, b) \subset \mathbb{R}$ and $u(x) = x$. Suppose $S = \{s_1, s_2\}$ with prior $\mu = (1/2, 1/2)$. Consider an act $f = (y, x)$, where $x \succ y$. Indifference curves for $\succsim$ and $\succsim_f$ are illustrated in Figure 1, where the dotted (black) curve represents $\succsim$ and the solid (blue) curve represents $\succsim_f$. ◊

There are several features of Figure 1 worth discussion. First, the endowment f causes the indifference curve to twist relative to the forty-five degree line (constant acts). This is because the slope of an indifference curve in utility space is completely determined by the relative probabilities of the states. Second, there are two shaded regions.

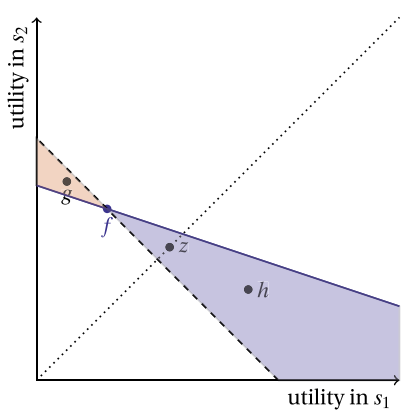

Figure 1. Indifference curves for $\succsim$ and $\succsim_f$.

The small region on the top left specifies all of the acts g which are worse than f according to the neutral preference $\succsim$ but are preferred to f according to $\succsim_f$, that is, $f \succ g$ and $g \succ_f f$. Note that every such g is more dispersed than f, i.e., $g(s_2) \succsim f(s_2) \succ f(s_1) \succsim g(s_1)$, which may result in the agent choosing something that is riskier than he otherwise would have. This is in line with experimental findings from Sprenger (2015) and Dean et al. (2017), in which endowing an agent with a lottery increases risk taking. The larger region on the bottom right specifies all acts h which are better than f according to the neutral preference $\succsim$ but are worse according to $\succsim_f$, i.e., $h \succ f$ and $f \succ_f h$.

To clarify, the agent does not privilege the endowment *per se*. Instead, he privileges those states of the world in which the endowment does well, perhaps because f makes such states more salient or vivid. Consequently, the agent may in fact choose something quite different from f. This serves as a distinction from other models of status quo bias, since f alters preferences rather than induces a mental constraint. By shifting probability mass to certain states, the agent may take "riskier" actions than he would in a neutral context, since he is more convinced that certain states will be realized. Put into a dynamic context, this unique interaction may result in a ratcheting effect, whereby an agent's beliefs induce more extreme actions which further distort beliefs, potentially contributing to extremism, polarization and escalation of commitment (Staw 1976).

Example 2 (Generalized status quo bias). Suppose $S = \{s_1, s_2\}$, $X = (0, 1)$, $\mu = (\frac{1}{2}, \frac{1}{2})$, and $u(x) = x$. Let $f = (x + \epsilon, x - \epsilon)$ and $f' = (x - \epsilon, x + \epsilon)$ for some $\frac{1}{2} \succ x + \epsilon \succ x \succ \epsilon$, and suppose δ_f and $\delta_{f'}$ are strictly increasing. Then let $h = (y, z)$ and $g = (z, y)$ for $y \succ z \succ \frac{1}{2}$. Absent a status quo, the agent ranks the above acts as follows: $h \sim g \succ f \sim f'$. When f is the status quo, $\mu_f(s_1) > \frac{1}{2} > \mu_f(s_2)$, and hence acts are ranked as follows: $h \succ_f g \succ_f f \succ_f f'$. Similarly, when f' is the status quo, acts are ranked as $g \succ_{f'} h \succ_{f'} f' \succ_{f'} f$. $\Diamond$

This is a more general version of the example from the introduction and preference reversals of this form have been observed in many economic experiments. However, they are impossible in almost all existing models of status quo bias. While maintaining a flavor of status quo bias, as $f \succ_f f'$ and $f' \succ_{f'} f$ will yield choice persistence, the reversals "above" the status quo are inconsistent with, for instance, Masatlioglu and Ok (2014). In particular, h and g are both ranked above f and f', yet their relative ranking changes across endowments. In contrast, this form of reversal is a robust prediction of the wishful thinking model.

4. Consequential distortion

The general representation allows beliefs across endowments to vary substantially. In particular, as long as the monotonicity property of δ_f is satisfied, the relative magnitude of the distortion may depend on the entire payoff profile of f. In this section, I introduce a restriction on behavior across endowments so that endowments which are similar on some states induce similar distortions. Hence I characterize the first special case by introducing an independence axiom which imposes that the distortion between any two states may only depend on the relative payoffs between those states, as defined below.

Definition 2 (Consequential distortion). The collection of Wishful Thinking Preferences, $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$, admits a *Consequential Distortion* if there exists an increasing function $v : u(X) \to \mathbb{R}_+$ such that for all $f \in \mathcal{F}$,

$$\delta_f(a) := \frac{v(a)}{\sum_{s\in S} v\big(u\big(f(s)\big)\big)\mu(s)}.$$

In this case, say that δ_f is a Consequential Distortion.

In this case, the belief distortion only depends on the endowment up to a normalizing constant; v is independent of f. To understand the intuition for this, consider any $A \subsetneq S$, and suppose $f(s) \sim g(s)$ for all $s \in A$. Without loss of generality, pick two acts, h and h', which provide 0 utility outside A. Since h and h' are identical outside A, the only thing that matters for comparison is their performance in A. Suppose $h \succsim_f h'$. How then might the agent rank h and h' in the event he had been endowed with g instead? It is sensible to think that however f impacts the agent's beliefs about states in A, g must impact them identically as f and g are identical on A. Therefore, it is reasonable to conclude that $h \succsim_g h'$ as well. The following axiom, Independence of irrelevant payoffs, imposes precisely the intuition from this example: for any two endowments, f and g, if there is an event on which they are payoff equivalent, then the preference ordering between any acts that vary only on that event must be the same across endowments.

Axiom 3 (Independence of irrelevant payoffs). *For all $f, g \in \mathcal{F}$ and any $A \subset S$, if $f(s) \sim g(s)$ for all $s \in A$, then for all $h, j \in \mathcal{F}$ and any $z \in X$,*

$$hAz \succsim_f jAz \iff hAz \succsim_g jAz.$$

Theorem 2. *Suppose the collection $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ admits a Wishful Thinking Representation with distortions $\{\delta_f\}_{f\in\mathcal{F}}$. Then the following are equivalent:*

(i) $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ satisfies Independence of irrelevant payoffs.

(ii) δ_f is a Consequential distortion.

Corollary 2. *Suppose the collection $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ admits a Consequential distortion. Then v is unique up to a positive scalar.*

While Corollary 1 shows the uniqueness of δ_f, the same uniqueness does not extend to the value function determining a Consequential distortion. That is, v is only identified up to the ratio of $\delta_f(a)$ and $\delta_f(b)$, as shown in Corollary 2.

Example 3 (Threshold distortion). Suppose for some thresholds, $\theta_L < \theta_H$, and strictly increasing real functions γ_L and γ_H,

$$v(a) = \begin{cases} \gamma_H(a) & a > \theta_H \\ 1 & \theta_L \le a \le \theta_H \\ \gamma_L(a) & a < \theta_L. \end{cases}$$

This general form nests many useful specifications. For instance, we might suppose for some $\lambda \in [0, \infty)$, $\gamma_j(a) = e^{\lambda(a-\theta_j)}$ for $j \in \{L, H\}$. $\diamondsuit$

For this specification, when possible payoffs are "moderate," $a \in [\theta_L, \theta_H]$, there is no distortion of beliefs. Hence probabilities are only distorted in the face of extreme outcomes: there is a possibility of something especially good, $a > \theta_H$, or especially bad, $a < \theta_L$.

When $\theta_L = \theta_H$, then there is always a distortion of probabilities. Further, in the case of $\gamma_j(a) = e^{\lambda(a-\theta_j)}$, the distortion reduces to $v(a) = e^{\lambda a}$, since v is only identified up to the ratio of $v(a)/v(b)$. This particular distortion was studied in Mayraz (2011a) (see Section 9.2) and can easily be used in applications, as shown in Section 8.

4.1 *Implications of Independence of irrelevant payoffs*

In this section, I provide an alternative characterization of the Consequential distortion using two conditions that are jointly weaker than Wishful Thinking. That is, in the presence of Independence of irrelevant payoffs, Wishful Thinking may be significantly weakened by replacing it with two new axioms. Both of the new axioms are implied by Wishful Thinking but do not imply Wishful Thinking when Independence of irrelevant payoffs is absent.

Recall that Wishful Thinking ensures the existence of a δ_f such that for any $x, y \in X$, (i) If $x \sim y$, then $\delta_f(u(x)) = \delta_f(u(y))$, and (ii) If $x \succ y$, then $\delta_f(u(x)) \geq \delta_f(u(y))$. The following axiom retains the first property, that equally good alternatives are equally distorted, while dropping the second property, monotonicity of δ_f.

Axiom 4 (Similar state consistency). *For all $f \in \mathcal{F}$ and $A \subset S$, if $f(s) \sim f(\tilde{s})$ for all $s, \tilde{s} \in A$, then for any $h, g \in \mathcal{F}$ and $z \in X$,*

$$hAz \succsim gAz \quad \Longleftrightarrow \quad hAz \succsim_f gAz.$$

Similar state consistency requires that for all events in which f yields a constant payoff, the ranking of acts that differ only on that event are consistent with the reference-free ranking. This places no restrictions on how states with different payoffs are distorted. In particular, Similar state consistency is consistent with a non-monotonic or even a strictly decreasing distortion.

The next axiom is a significant weakening of Wishful Thinking that simply requires that whenever an agent prefers f to some constant action x according to the neutral preference, then f must still be preferred when f is the status quo.

Axiom 5 (Minimal status quo bias). *For all $f, x \in \mathcal{F}$,*

$$f \succsim x \quad \Rightarrow \quad f \succsim_f x.$$

As this axiom is implied by Weak Status Quo Bias from Masatlioglu and Ok (2014), which is further discussed in Section 9.1, I refer to it as Minimal Status Quo Bias.

When combined with Consistent expected utility, Minimal status quo bias only ensures that $V_f(f) \geq V(f)$. In terms of its implications for beliefs, δ_f may be non-monotonic, though it must be increasing for binary acts. That is, for any $x, y \in X$, and $A \subset S$, if $x \succ y$ then $\delta_{xAy}(u(x)) \geq \delta_{xAy}(u(y))$. Hence it adds back some limited monotonicity of the belief distortion. However, by combining this result with Independence of irrelevant payoffs, v can be constructed from binary acts and then extended to all f.

Theorem 3. *Suppose the collection $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ satisfies Consistent expected utility and Independence of irrelevant payoffs. Then the following are equivalent:*

(i) $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ satisfies Wishful Thinking.

(ii) $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ satisfies Minimal status quo bias and Similar state consistency.

5. Binary distortions

There are many interesting and intuitive examples of distortions that depend on the endowment in a more general way. For example, the distortion may take the form of a step function (similar to Example 3) in which the agent separates S into *good* and *bad* states. It is natural for the notion of good and bad, the thresholds θ_2 and θ_1, to depend on f. This section introduces and characterizes a special case of the class of binary distortions in which *good* states are those in which the endowment yields its highest possible payoff.

Definition 3. For any $f \in \mathcal{F}$, $\mathcal{D}(f) = \{s \in S | f(s) \succsim f(s') \text{ for all } s' \in S\}$.

Thus $\mathcal{D}(f)$ is the set of f-optimal states; those in which f yields its maximal payoff:

Definition 4 (Best-case binary distortion). The collection of Wishful Thinking Preferences, $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ admits a *Best-Case Binary Distortion*[7] if there is a function $\delta : \mathcal{F} \to [0, 1]$ such that

$$\mu_f(s) = \big(1 - \delta(f)\big)\mu(s) + \delta(f)\mu\big(s|\mathcal{D}(f)\big).$$

For every f the agent partitions S into *good* and *bad* states, where good states are those in which f attains its maximal payoff and beliefs move in the direction of the good states. The size of this movement is given by $\delta(f)$. In order to characterize the Best-case binary distortion, I will utilize a strong notion of "aligned" acts, defined as follows.

Definition 5 (Strong comonotonicity). Say that h and f are *strongly comonotonic* if $h(s) \succsim h(s')$ if and only if $f(s) \succsim f(s')$ for all $s, s' \in S$. Denote this by $h \asymp f$.

[7] This representation fits into the general model as follows: given a particular, a binary distortion $\delta : \mathcal{F} \to [0, 1]$, we define $t : \mathcal{F} \to \mathbb{R}$, as $t(f) = \max_{s\in S} u(f(s))$, and then for each $a \in \mathbb{R}$,

$$\delta_f(a) := \begin{cases} 1 - \delta(f) & \text{if } a < t(f) \\ 1 - \delta(f) + \delta(f)\dfrac{1}{\mu\big(\{s \in S | u\big(f(s)\big) \geq t(f)\}\big)} & \text{if } a \geq t(f). \end{cases}$$

Consider comparing h and g when endowed with f, and suppose $h \asymp g \asymp f$. If (i) $h \succsim g$, then a standard agent ($\delta(f) = 0$) must prefer h to g when endowed with f. If (ii) $h(s) \succsim g(s')$, for some $s \in S$ and all $s' \in S$, then a maximally wishful ($\delta(f) = 1$) agent will prefer h to g when endowed with f. If both (i) and (ii) are true, h must be preferred to g for any level of wishful thinking.

Axiom 6 (Best-case dominance). *For any $f, g, h \in \mathcal{F}$, if $h \asymp g \asymp f$, then*

$$\left.\begin{array}{l} h \succsim g \\ h(s) \succsim g(s'), \text{ for some } s \in S \text{ and all } s' \in S \end{array}\right\} \quad \Longrightarrow \quad h \succsim_f g.$$

Best-case dominance places substantial structure on the connection between neutral and conditional preferences. In particular, it is insufficient to just know that h is preferred to g in the neutral ranking. But when h is strongly comonotonic with f and the maximal payoff of h is better than the maximal payoff of g, then the neutral ranking must be preserved.

Theorem 4. *Suppose the collection $\{\succsim, \succsim_f\}_{f \in \mathcal{F}}$ admits a Wishful Thinking Representation. Then the following are equivalent:*

(i) $\{\succsim, \succsim_f\}_{f \in \mathcal{F}}$ satisfy Best-case dominance.

(ii) $\{\succsim, \succsim_f\}_{f \in \mathcal{F}}$ admits a Best-case binary distortion.

Corollary 3. *Suppose the collection $\{\succsim, \succsim_f\}_{f \in \mathcal{F}}$ admits a Best-case binary distortion. Then $\delta(f)$ is unique for all $f \in \mathcal{F}$ such that $f(s) \succ f(\tilde{s})$ for some $s, \tilde{s} \in S$.*

It is important to note that while the distortion is restricted to only overweight states in $\mathcal{D}(f)$, the weight on those states, $\delta(f)$, may depend on f more generally.

6. Connecting the two cases

Both of the special cases presented so far may be useful formulations of wishful thinking in certain contexts. The Consequential distortion (CD) allows for belief distortions to be partially independent of the endowment, while the Best-case binary distortion (BCB) captures a simple cognitive mechanism and allows for natural forms of endowment dependence. This section explores the relation between the two cases by imposing both Independence of irrelevant payoffs and Best-case dominance. Both special cases are essentially disjoint: an agent may satisfy both conditions only if the agent exhibits no wishful thinking. The content of Theorem 5 is illustrated in the upper portion of Figure 2.

Theorem 5. *Suppose $\{\succsim, \succsim_f\}_{f \in \mathcal{F}}$ admits a Wishful Thinking Representation. Then the following are equivalent*

(i) $\{\succsim, \succsim_f\}_{f \in \mathcal{F}}$ satisfy Independence of irrelevant payoffs and Best-case dominance.

(ii) $\succsim = \succsim_f$ for all $f \in \mathcal{F}$.

Proof. (i) $\Rightarrow$ (ii) Consider consequences $w \succ x \succ y \succ z$. Suppose $\{\succsim, \succsim_f\}$ satisfies both Independence of irrelevant payoffs and Best-case dominance. Then fix some $s \in S$ and consider $f = y\{s\}z$ and $g = w\{s\}x$. For the purpose of showing a contradiction, suppose $\succsim_f \neq \succsim_g \neq \succsim$. Then $\frac{\mu_f(s)}{\mu_f(s')} = \frac{v(u(y))}{v(u(z))}\frac{\mu(s)}{\mu(s')} = \frac{(1-\delta(f))\mu(s)+\delta(f)}{(1-\delta(f))\mu(s')}$ and $\frac{\mu_g(s)}{\mu_g(s')} = \frac{v(u(w))}{v(u(x))}\frac{\mu(s)}{\mu(s')} = \frac{(1-\delta(g))\mu(s)+\delta(g)}{(1-\delta(g))\mu(s')}$. By assumption, $\delta(f) > 0$ and $\delta(g) > 0$, hence $v(u(w)) > 1 > v(u(x))$ and $v(u(y)) > 1 > v(u(z))$. However, this implies $v(u(y)) > v(u(x))$, a contradiction of monotonicity of v. (ii) $\Rightarrow$ (i) Since $\succsim = \succsim_f$ is clearly a special case of both models, it must be that $\{\succsim, \succsim_f\}$ satisfies both Independence of irrelevant payoffs and Best-case dominance. □

7. Comparative wishful thinking

This section develops a comparative notion of wishful thinking by providing a behavioral definition of when one agent is *more wishful* than another. Consider two agents that satisfy the conditions of Theorem 1. For $i = 1, 2$, let $\{\succsim^i, \succsim^i_f\}_{f\in\mathcal{F}}$ denote i's preferences. The following definition is similar in spirit to definitions of *more ambiguity averse* or *more status quo biased*.

Definition 6. Given two agents, agent two is *more wishful* than agent one if $\succsim^2 = \succsim^1$ and for all $f \in \mathcal{F}$,

$$f \succsim^1_f x \quad \Rightarrow \quad f \succsim^2_f x.$$

Neutral preferences are assumed to be identical in order to control for beliefs and tastes. Then agent two is more wishful if she always values the endowment more than agent one. The following result relates this preference-based definition of *more wishful* to the model parameters for the Consequential distortion and Best-case binary distortion special cases.

Theorem 6. *Suppose agents 1 and 2 admit a Wishful Thinking Representation. Then:*

(i) If both agents admit a Consequential distortion, agent two is more wishful than agent one if and only if for all $a, b \in u(X)$, $a \geq b$,

$$\frac{v^2(a)}{v^2(b)} \geq \frac{v^1(a)}{v^1(b)}.$$

(ii) If both agents admit a Best-case binary distortion representation, agent two is more wishful than agent one if and only if for all $f \in \mathcal{F}$,

$$\delta^2(f) \geq \delta^1(f).$$

8. Applications

8.1 *Asset markets and wishful thinking*

The first application illustrates the effects of wishful thinking in a market setting. Agents are endowed with Arrow securities in an economy with no aggregate risk and are allowed to trade. The main finding is that while standard agents fully diversify in this

setting, the existence of any bias due to wishful thinking results in under-diversification and therefore equilibria are typically not Pareto efficient (relative to the no-endowment preference $\succsim$).

Formally, there are two states, $S = \{A, B\}$ and two Arrow securities, a and b, which are in unit supply and pay one unit of the consumption commodity in states A and B, respectively. There are two agents, 1 and 2, with consumption utility $\ln(c)$. At time 0, agents realize their endowments, $\omega_i = (\omega_{ia}, \omega_{ib}) \gg (0, 0)$, and trade. At time 1, the state is realized and consumption occurs. This setting may be embedded into the framework of this paper as follows: $X = (-\infty, 0)$ and each allocation, (a_i, b_i) is identified with the act $f_{(a_i,b_i)} = (\ln(a_i), \ln(b_i))$. Note that $a_i = b_i = x$ corresponds to the constant act $f(j) = x$ for $j \in \{A, B\}$. Hence an endowment ω_i and prices $p = (p_a, p_b)$ induce preferences $\succsim^i_{f_{(\omega_{ia},\omega_{ib})}}$ and choice set $\mathcal{B}(\omega_i, p) = \{f_{(a_i,b_i)} | p_a a_i + p_b b_i \leq p_a \omega_{ia} + p_b \omega_{ib}\}$. Through the rest of this section, I will simplify notation by simply referring to allocations, rather than their induced acts.

An economy is given by $((\omega_1, \omega_2), (\succsim^1_{\omega_1}, \succsim^2_{\omega_2}))$, where $\succsim^i_{\omega_i}$ admits a Wishful Thinking Representation for each i. The key departure from the standard model is that endowments determine choices not just via the budget constraint but through their impact on preferences by directly influencing beliefs over the states. Hence I will also suppose that agents' endowment-free preferences are identical, $\succsim^1 = \succsim^2$, and thus results will be driven by the effect of endowments on beliefs rather than fundamentally different priors or tastes.[8] Let $\mu(A)$ denote the reference-free probability for state A and $w_i = p_a \omega_{ia} + p_b \omega_{ib}$ stand for i's induced wealth.

Definition 7 (Equilibrium). A price $p^* = (p^*_a, p^*_b) \in \mathbb{R}^2$ and allocations $((a_1, b_1), (a_2, b_2)) \in [0, 1]^4$ constitute a competitive equilibrium if:

(i) $(a_1, b_1) + (a_2, b_2) = (1, 1)$.

(ii) $(a_i, b_i) \succsim^i_{\omega_i} (a'_i, b'_i)$ for all $(a'_i, b'_i) \in \mathcal{B}(\omega_i, p^*)$, for each i.

The definition of competitive equilibrium is standard, consisting of prices and allocations such that (i) markets clear and (ii) agents' allocations are preference maximal given prices and endowments. However, implicit in this equilibrium is the assumption that final allocations only depend on the initial endowment. This can be thought of as either a weak form of naïveté or the assumption that belief changes are "slow" and there is no possibility of retrading. Beliefs are solely determined by the exogenous endowment, not possible future holdings.[9]

To clarify the role of wishful thinking, I will make a few additional assumptions on initial endowments and belief distortions. Suppose the agents are identical except for

[8]That is, agents are *ex ante* identical and are only different upon realization of their endowments.

[9]In principle, agents could trade and then have an instantaneous change in beliefs, reflecting their new holdings, which would induce a desire to re-trade. The possible implications of this are partially discussed later in this section. One way around this would be to build in this iterative belief change to the definition of equilibrium and assume something akin to personal equilibrium: $(a_i, b_i) \succsim^i_{(a_i,b_i)} (a'_i, b'_i)$ for all $(a'_i, b'_i) \in \mathcal{B}((a_i, b_i), p^*)$, for each i.

the relative distribution of endowments. That is, suppose preferences conditional on the same endowment are identical ($\succsim_f^1 = \succsim_f^2$), that states are equally likely *ex ante* ($\mu(A) = \mu(B)$), and that endowments are symmetric ($\omega_{1a} = \omega_{2b}$). Further, suppose that each agent admits a Consequential distortion where $u : X \to \mathbb{R}$ and $v : u(X) \to \mathbb{R}_{++}$ are given by $u(x) = x$ and $v(\tilde{u}) = e^{\alpha\tilde{u}}$. Hence, $\mu_{\omega_i}(A) = \frac{\omega_{ia}^\alpha}{\omega_{ia}^\alpha + \omega_{ib}^\alpha}$.[10]

Proposition 1. *Suppose $\alpha_1 = \alpha_2 = \alpha$. Then for every $\alpha \geq 0$, a unique equilibrium exists with $p^* = (\frac{1}{2}, \frac{1}{2})$ and*

$$(a_i^*, b_i^*) = \left(\frac{\omega_{ia}^\alpha}{\omega_{ia}^\alpha + \omega_{ib}^\alpha}, \frac{\omega_{ib}^\alpha}{\omega_{ia}^\alpha + \omega_{ib}^\alpha} \right).$$

Further,

(i) If $\alpha = 0$, then there is no wishful thinking (each agent is standard) and full risk sharing: $(a_i^*, b_i^*) = (\frac{1}{2}, \frac{1}{2})$.

(ii) If $0 < \alpha < 1$, then there is partial risk sharing.

(iii) If $\alpha = 1$, then there is no trade.

(iv) If $\alpha > 1$, then agents take on additional risk.

This result highlights a novel effect of inequality in the market. Even though agents are equal in expected wealth, their endowments differ in their wealth distribution across states. This skews their beliefs toward their endowment and results in inefficient risk sharing, relative to the endowment-free preference. Note that this is the case even in the face of "correct" prices and illustrates that only observing prices may not be sufficient to conclude that the economy has reached an efficient outcome.

When agents are more wishful (α is larger), their beliefs are more skewed and they hold more risk than they would without an endowment. Proposition 1 also illustrates the importance of the assumption on when beliefs change. In particular, if there is a lag in belief change or agents are naive, then the equilibrium definition used is likely the correct notion. However, if we assume instantaneous belief change or sophistication, then the possibility of retrade may change the equilibrium predictions. To see how, notice that when $\alpha < 1$, agents engage in partial risk sharing, and thus they trade away from their endowment toward $(\frac{1}{2}, \frac{1}{2})$, though not fully. If belief adjustment is instant, their new asset holdings will push their beliefs closer to μ and induce further risk sharing. Similarly, for $\alpha > 1$, agents take on additional risk. This will drive beliefs apart and induce a ratchet effect, where agents repeatedly re-trade toward more extreme holdings. This is not empirically plausible, and so either adjustment of beliefs is slow as suggested or there is some other mechanism limiting this. I will briefly discuss one possibility: the case where only agent one is biased, $\alpha_1 > 0$, and agent two is standard, $\alpha_2 = 0$. Without loss of generality, suppose $\omega_{1a} > \omega_{1b}$.

[10]Existence of equilibrium does not rely on the assumption of a Consequential distortion and can be shown for general wishful thinking preferences.

PROPOSITION 2. *An equilibrium exists and*

$$\frac{p_a^*}{p_b^*} = \frac{1+\omega_{1b}\left(\dfrac{\omega_{1a}^\alpha - \omega_{1b}^\alpha}{\omega_{1a}^\alpha + \omega_{1b}^\alpha}\right)}{1-\omega_{1a}\left(\dfrac{\omega_{1a}^\alpha - \omega_{1b}^\alpha}{\omega_{1a}^\alpha + \omega_{1b}^\alpha}\right)}.$$

Notice that in the limit, as α increases to infinity, $\frac{p_a^*}{p_b^*} = \frac{1+\omega_{1b}}{1-\omega_{1a}}$. Thus the wishful thinking agent may influence prices, but has limited power to do so. Hence even if the wishful agent's beliefs were to "drift off," prices converge, and hence so will allocations.

8.2 *Willingness to accept and willingness to pay*

The second application derives the implications of wishful thinking on the well-known gap between willingness to accept (*wta*) and willingness to pay (*wtp*) for endowed goods.[11] The first major implication is that the model predicts a gap in most instances. The second major implication is that the magnitude of the gap may be context dependent; the size of the gap depends on the distribution of utility across states. I will consider a general setting rather than one closely aligned with any particular experiment. Suppose $S = \{s_1, \dots, s_n\}$ and $X = [0, \infty)$ so that acts can be interpreted as having monetary payoffs. In this case, *wtp* and *wta* can be easily determined by performing the right utility evaluation. In particular, for any $f \in \mathcal{F}$

(i) $wta(f) := \inf\{x \geq 0 | x \succsim_f f\} = V_f(f) = \sum_{i=1}^n f(s_i)\mu(s_i)\delta_f(f(s_i))$.

(ii) $wtp(f) := \sup\{x \geq 0 | f \succsim x\} = V(f) = \sum_{i=1}^n f(s_i)\mu(s_i)$.

PROPOSITION 3. *For any $f \in \mathcal{F}$, $wta(f) - wtp(f)$ is given by*

$$V_f(f) - V(f) = \sum_{i=1}^n f(s_i)\mu(s_i)\big[\delta_f\big(f(s_i)\big) - 1\big] \geq 0.$$

PROOF. To see this, recall that δ_f is increasing for every f, and thus for any two states, $f(s_i) \succsim f(s_j)$ implies $\frac{\mu_f(s_i)}{\mu_f(s_j)} = \frac{\delta_f(f(s_i))\mu(s_i)}{\delta_f(f(s_j))\mu(s_j)} \geq \frac{\mu(s_i)}{\mu(s_j)}$. Given this property, the implied distribution over payoffs of f under μ_f first-order stochastically dominates the distribution of payoffs under μ, hence $V_f(f) \geq V(f)$. □

There are three key insights from Proposition 3. First, the *wta–wtp* gap is always weakly positive, though its magnitude depends on the nature of the object. Second, since $V_x(x) = V(x) = u(x)$ for all certain acts, $wta(x) - wtp(x) = 0$, and thus wishful thinking only induces a *wta–wtp* gap when there is uncertainty.[12] Third, the *wta–wtp*

[11] The classic mug experiment is from Knetsch (1989).

[12] This is in line with experimental findings of Plott and Zeiler (2005), Plott and Zeiler (2011), and Isoni et al. (2011), which found no gap for mugs (certain objects) but did find gaps for lotteries (uncertain objects). Further, Plott and Zeiler (2011) found that the evidence suggests the observed gaps are due to shifting of beliefs, in which lottery holders (purchasers) put too much (too little) weight on the high payoff, consistent with wishful thinking.

gap may depend on the entire distribution of payoffs, and thus an agent susceptible to wishful thinking may be more reluctant to give up objects with a chance at very high payoffs. For illustration, consider the following example.

Example 4. Suppose preferences admit a Consequential distortion and $S = \{s_1, s_2\}$. In this case, whenever f' is a mean-preserving spread of f, then $wta(f') - wtp(f') \geq wta(f) - wtp(f)$. The logic behind this is simple. Since f' is a mean-preserving spread of f, it is without loss to suppose $f'(s_1) \succ f(s_1) \succ f(s_2) \succ f'(s_2)$. Since v is increasing this ensures that $\mu_{f'}(s_1) > \mu_f(s_1)$. When combined with $V(f') = V(f)$, it is straightforward to see that $wta(f') - wtp(f') > wta(f) - wtp(f)$. $\Diamond$

Lastly, it is worth mentioning that the *wta–wtp* gap takes a particularly nice form when preferences admit a Best-case binary distortion. In particular,

$$wta(f) - wtp(f) = \delta(f)\Big(\max_{s\in S} u\big(f(s)\big) - \sum_{s\in S} u\big(f(s)\big)\mu(s)\Big).$$

Thus the gap only depends on the magnitude of the belief distortion, $\delta(f)$, and the maximal utility of the endowed act.

9. Relation to other models

9.1 *Comparison with Masatlioglu and Ok (2014)*

This section formally analyzes the relation to the model of status quo bias à la Masatlioglu and Ok (2014) (which generalizes Masatlioglu and Ok 2005).[13] Adapting their model to the current setting, a choice problem is a pair (M, σ) where M is a finite set of acts, $M \subset \mathcal{F}$, and $\sigma \in \mathcal{F} \cup \{\diamond\}$, where $\sigma = \diamond$ represents choice without status quo or neutral context. Their primitive is a choice correspondence from the collection of choice problems satisfying $C(M, \sigma) \subset M$ for every σ. Within this setting, their key axioms are below.

Axiom 7 (Weak status quo bias). *For any $f, g \in \mathcal{F}$,*

(i) $g \in C(\{f, g\}, f) \Rightarrow g \in C(\{f, g\}, \diamond)$.

(ii) $g \in C(\{f, g\}, \diamond) \Rightarrow g \in C(\{f, g\}, g)$.

Axiom 8 (Status quo irrelevance). *For any (M, f), suppose $\{f\} \neq C(N, f)$ for every non-singleton subset N of M with $f \in N$ and that $g = C(\{f, g\}, \diamond)$ for some $g \in M$. Then $C(M, f) = C(M, \diamond)$.*

They show that the above axioms, along with the weak axiom of revealed preference with a fixed endowment and continuity, ensure the following representation: There exists a continuous $U : \mathcal{F} \to \mathbb{R}$, and a closed valued correspondence $\mathcal{Q}$ on $\mathcal{F}$ such that $C(M, \diamond) = \arg\max U(M)$ and $C(M, f) = \arg\max U(M \cap \mathcal{Q}(f))$.

[13]Since the analysis is done at the axiomatic level, the conclusions apply to all models utilizing their axioms, WSQB or SQI, such as Ortoleva (2010) and Riella and Teper (2014).

In order to facilitate a more direct comparison of their model with the model here, define the wishful thinking choice correspondence as $C_{\mathrm{WT}}(M,\diamond) = \{g \in M | g \succsim h \text{ for all } h \in M\}$ and $C_{\mathrm{WT}}(M,f) = \{g \in M | g \succsim_f h \text{ for all } h \in M\}$. The following two examples demonstrate that the model of wishful thinking typically violates both Weak status quo bias (WSQB) and Status quo irrelevance (SQI).

Example 5 (Violation of WSQB). Let $S = \{s_1, s_2\}$, $X = [0,1]$, and $f = (\frac{3}{4}, \frac{1}{4})$. $v(\tilde{u}) = \tilde{u}$ and $u(x) = x$. Then $\mu = (\frac{1}{2}, \frac{1}{2})$ and $\mu_f = (\frac{3}{4}, \frac{1}{4})$. Let $g = (1-\epsilon, 0)$. Then for $0 < \epsilon < \frac{1}{6}$, $g = C_{\mathrm{WT}}(\{f,g\}, f)$ and $f = C_{\mathrm{WT}}(\{f,g\},\diamond)$, which violates the first part. Notice, for $\epsilon = 0$, $f \in C_{\mathrm{WT}}(\{f,g\},\diamond)$ but $C_{\mathrm{WT}}(\{f,g\},f) = g$, which violates the second part. $\Diamond$

Example 6 (Violation of SQI). Using the same setup as the previous example, $M = \{f, g, x_1, \dots, x_n\}$, where x_i is a constant act with $x_i(s_j) \in (\frac{5}{8}, \frac{3(1-\epsilon)}{4})$ for each i, $x_i < x_{i+1}$ and $\epsilon \in (0, \frac{1}{6})$. Then for any nonsingleton $N \subset M$, $f \notin C_{\mathrm{WT}}(N,f)$, $C_{\mathrm{WT}}(M,f) = g$, but $C_{\mathrm{WT}}(M,\diamond) = x_n$. $\Diamond$

The reason for these violations is intuitive. In Masatlioglu and Ok (2014), the status quo exerts a pull on the agent without otherwise distorting preferences. The strength of this pull is captured by $\mathcal{Q}$. In particular, the utility for a given outcome is status quo-independent. In contrast, wishful thinking actually modifies the agent's preferences by twisting his indifference curves in utility space. Thus even when the status quo is not chosen the status quo is never irrelevant because it directly influences beliefs over the states.

Further, these examples are not just carefully crafted instances but reflect a more fundamental distinction between the two models. In fact, imposing either Weak status quo bias or Status quo irrelevance on wishful thinking preferences completely rules out all impact of the status quo (see Figure 2). To see this, consider the preference formulations of the MO axioms. Weak status quo bias states that for any $f, g, h \in \mathcal{F}$, (i) $g \succsim_f f \Rightarrow g \succsim f$ and (ii) $g \succsim f \Rightarrow g \succsim_g f$. Status quo irrelevance states that for any $f, g, h \in \mathcal{F}$ such that $g, h \succsim_f f$ and $g, h \neq f$, then $g \succsim_f h \Leftrightarrow g \succsim h$. I also need the following technical property on X: say X *has no upper bound* if for all $x \in X$, there exists $z \in X$ with $z \succ x$.

Theorem 7. *Suppose the collection $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ admits a Wishful Thinking Representation. Then,*

(i) If $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ satisfies Weak status quo bias and X has no upper bound, then $\succsim = \succsim_f$ for all $f \in \mathcal{F}$.

(ii) If $\{\succsim, \succsim_f\}_{f\in\mathcal{F}}$ satisfies Status quo irrelevance, then $\succsim = \succsim_f$ for all $f \in \mathcal{F}$.

Proof. (i) Suppose $u(X) \supset [0,\infty)$ and suppose for sake of contradiction that $\succsim_f \neq \succsim$ for some f. Decompose S into a partition, $\{E_1, \dots, E_n\}$, where for each i, f is a constant and for $i < j$, and $s_k \in E_k$, $f(s_i) \succ f(s_j)$. Then necessarily $\delta_f(E_1) > 1$ and hence $\mu_f(E_1) > \mu(E_1)$. Construct g as follows. Pick any number $\gamma_1 > 0$ and for all $s \in S \backslash E_1$,

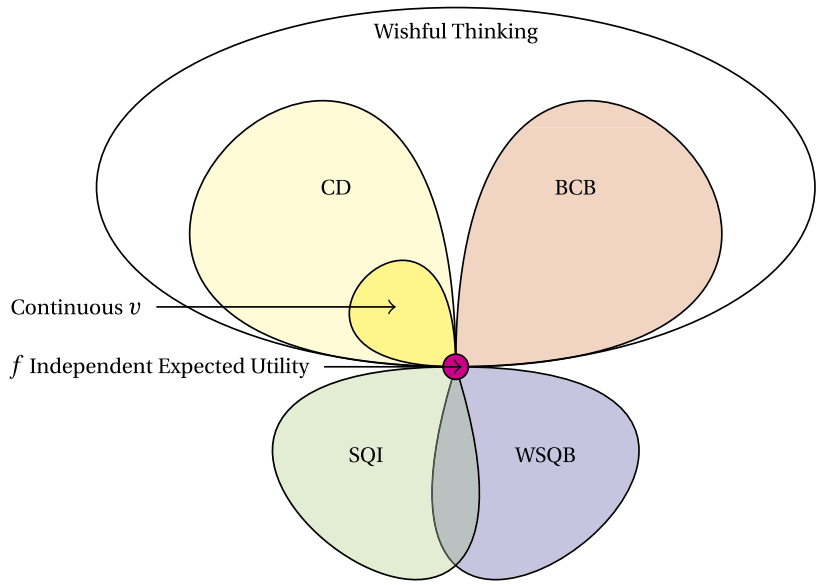

FIGURE 2. Illustration of model relationships.

let $g(s)$ satisfy $u(f(s)) - u(g(s)) = \gamma_1$. Next, let $\gamma_2 = \gamma_1 \frac{1-\mu_f(E_1)}{\mu_f(E_1)} > 0$ and for $s \in E_1$, let $g(s)$ satisfy $u(g(s)) - u(f(s)) = \gamma_2$. By construction, $V_f(g) - V_f(f) = \gamma_2 \mu_f(E_1) - \gamma_1(1 - \mu_f(E_1)) = 0$, and thus $g \sim_f f$. In addition, $V(g) - V(f) = \gamma_2 \mu(E_1) - \gamma_1(1 - \mu_f(E)) = \gamma_1 \frac{1-\mu_f(E_1)}{\mu_f(E_1)} \mu(E_1) - \gamma_1(1 - \mu(E_1)) < 0$, and thus $f \succ g$, a violation of Weak status quo bias condition (i). The construction of a violation for condition (ii) is similar. The assumption that $u(X) \supset [0, \infty)$ is needed to avoid certain problematic cases. In particular, recall Figure 1 and note that the g constructed here lies in the upper left triangle; the existence of this region is necessary to construct this contradiction. If f is an interior act (and $\succsim \neq \succsim_f$), then this triangle will always exist. When $u(X) \supset [0, \infty)$, f is always interior.

(ii) Consider any $f \in \mathcal{F}$ and suppose $\succsim_f \neq \succsim$. Necessarily there exists s and s' such that $f(s) \succ f(s')$. Decompose S into a partition, $\{E_1, \dots, E_n\}$ and not that necessarily $\mu_f(E_1) > \mu(E_1)$. Pick $\gamma_1, \gamma_2, \gamma_3, \gamma_4 \in u(X)$, such that $\gamma_1 > \gamma_2$, $\gamma_3 > \gamma_4$, and $\frac{1-\mu(E_1)}{\mu_(E_1)} > \frac{\gamma_1 - \gamma_2}{\gamma_3 - \gamma_4} > \frac{1-\mu_f(E_1)}{\mu_f(E_1)}$. Then let $x_i = u^{-1}(\gamma_i)$ for each i and $h = x_1 E_1 x_4$ and $g = x_2 E_1 x_3$. It is without loss to assume $h, g \succ_f f$, since by independence we can mix h and g with something strictly preferred to f. Finally, it can be verified through calculation that $h \succ_f g$ and $g \succ h$. □

9.2 *Comparison with Mayraz (2011a)*

Mayraz (2011a) characterizes a nice model of payoff-dependent beliefs in which the probability ratio between two states is distorted by the exponential of the utility difference in those two states. His setting is slightly different, as there is no object which corresponds to the reference-free preference in this paper (which makes comparison

to status quo bias indirect). Nonetheless, we can compare his representation of conditional preferences to the representations in this paper. Formally, he characterizes the following (adapted to the framework of this paper): there is some $\lambda \in \mathbb{R}$, such that

$$\frac{\mu_f(s)}{\mu_f(\tilde{s})} = \frac{\mu(s)}{\mu(\tilde{s})} e^{\lambda[u(f(s))-u(f(\tilde{s}))]}.$$

When $\lambda \geq 0$, his model is a special case of the Consequential distortion (with continuous v, as illustrated in Figure 2) where $v(a) = e^{\lambda a}$. As his model allows for $\lambda < 0$, which corresponds to "pessimism" or the underweighting of good events, his model is not fully nested by the Wishful Thinking Representation.[14] The ability to elegantly capture both optimism and pessimism with a single parameter is a noted strength of his model. Since the model he characterizes does not require a particular direction of bias, his axiomatization necessarily reflects this and, therefore, does not contain anything which could be considered a "behavioral characterization" of wishful thinking (or optimism in his terminology) akin to Wishful Thinking or Minimal status quo bias.

Mayraz utilizes nine axioms, B1–B9, some of which are implied by, or correspond to, axioms in this paper. B1 ensures an expected utility representation. B2 is "ordinal preference consistency" and is implied by Similar state consistency (hence also Wishful Thinking), while B3 and B7 are implied by Independence of irrelevant payoffs. His other axioms are the following: B4, which is a nonindifference condition assumed in Consistent expected utility; B6, which ensures consistent treatment of null events (required for his more general state space); B8, which imposes "shift-invariance" and is essential to achieving the exponential functional form; and B9, a continuity condition.

In terms of behavior, Mayraz requires that the probability ratio of two states is distorted for all utility differences. The Consequential distortion, which generalizes his representation (under optimism) allows for an agent to distort only those states featuring "extreme payoffs," as discussed in Example 3.

9.3 *Comparison with Brunnermeier and Parker (2005)*

Since f distorts beliefs in an optimistic fashion, this suggests a close relation to Brunnermeier and Parker (2005) (BP), but there are some key differences. As was shown in Spiegler (2008), the BP choice correspondence may violate Independence of Irrelevant Alternatives (IIA). This is because beliefs and actions are chosen jointly and therefore the set of available actions influences the chosen beliefs. In contrast, wishful thinking beliefs are distorted by an exogenous endowment and are independent of the menu from which a future choice may be made. Given this, choices under wishful thinking will satisfy IIA for a fixed endowment, though choice reversals may occur across endowments.[15]

[14] However, it is straightforward to modify Wishful Thinking (or Minimal status quo bias) to obtain a decreasing distortion function which would cover negative λ.

[15] If instead I were to assume f is an endogenous reference corresponding to a personal equilibrium as in Kőszegi and Rabin (2006), then IIA may be violated. However, imposing personal equilibrium amounts to a different model, and so the exact connection between a wishful thinking-personal equilibrium and the choice model of BP, while interesting, is beyond the scope of this paper.

9.4 *Other related literature*

This paper bridges two literatures by linking status quo bias with belief biases. As both literatures are quite expansive, I will focus on the most closely related papers.

Status Quo Bias: Both Ortoleva (2010) and Riella and Teper (2014) consider choice over acts with a status quo. Ortoleva (2010) uses the status quo bias axiom from Masatlioglu and Ok (2005) and derives a unanimity representation as in Bewley (2002). Riella and Teper (2014) utilize the status quo irrelevance axiom from Masatlioglu and Ok (2005) and derive a representation where the decision maker is constrained to choose among alternatives that are very likely to be better than the status quo. In both papers, beliefs are independent from the status quo. Dean et al. (2017) combines limited attention with status quo bias from Masatlioglu and Ok (2014) to explain some forms of generalized status quo bias. In the "LA-SQB" model, however, choice is still determined by the maximization of a reference-free preference over some constraint set, hence the wishful thinking model and LA-SQB are generally distinct. In particular, wishful thinking choice will violate their Axiom 3. For other papers on status quo and reference dependence, see Rubinstein and Zhou (1999), Apesteguia and Ballester (2009), and Tserenjigmid (2019).

Belief biases: Yariv (2001) studies a generalization of the discounted utility model in which preferences are over pairs of actions and beliefs. Yariv (2005) considers an agent represented by an instrumental utility and a belief utility, where the belief utility captures the agent's innate preference for belief consistency. In each period, the agent's choice is over beliefs, subject to the constraint that the agent will take an action consistent with his beliefs and suffers a cost of changing beliefs (see also Bénabou and Tirole 2011, Bénabou 2013). Epstein and Kopylov (2007) study an agent who balances his preferences under commitment against a temptation utility. Caplin and Leahy (2001) consider a two-period model in which the agent's utility is defined over both prizes and psychological states and studies the role of anticipatory feelings. This agent's *ex ante* beliefs balance his instrumental utility with utility from anticipation (or anxiety). Lastly, Eyster (2002) considers a decision maker with a taste for consistency between actions, which arises due to feelings of regret when a past action is suboptimal. This taste for consistency only affects choice when there are multiple options available at each period. Hence, if the first-period choice set is a singleton, then choice is "standard." In the wishful thinking model (though the origin of the endowment is not part of the model), the endowment always affects choice.

Appendix A: Preliminary results

Axiom 1, Consistent expected utility, consists of the following postulates. For all $f, g, h, h' \in \mathcal{F}$, $\succsim$ and $\succsim_f$ satisfy (i) **Weak order**: they are complete and transitive binary relations, (ii) **Independence**: for all $\alpha \in (0,1)$, $h \succsim (\succsim_f) h'$ if and only if $\alpha h + (1-\alpha) g \succsim (\succsim_f) \alpha h' + (1-\alpha) g$, (iii) satisfy **Strict monotonicity**: If $h(s) \succsim (\succsim_f) g(s)$ for all s, then $h \succsim (\succsim_f) g$. In addition, if for some s, $h(s) \succ (\succ_f) g(s)$, then $h \succ (\succ_f) g$, (iv) **Continuity**: the weak upper and lower-contour sets are closed, and (v) **Nontriviality:** there are $x, y \in X$ such that $x \succ y$.

Lemma 1. *Under Consistent expected utility and Similar state consistency, the collection of preferences satisfy* ***Ordinal Preference Consistency:*** *For all $f \in \mathcal{F}$ and $x, y \in X$, $x \succsim y$ if and only if $x \succsim_f y$.*

Proof. First, suppose $x \succsim y$. Then equivalently, for any $s \in S$, $x\{s\}z \succsim y\{s\}z$, by strict monotonicity. Then for any f and taking $A = \{s\}$, by Similar state consistency $x\{s\}z \succsim y\{s\}z \Leftrightarrow x\{s\}z \succsim_f y\{s\}z$. However, the latter is equivalent to $x \succsim_f y$, again by strict monotonicity. □

Lemma 2. *There exists a nonconstant, affine utility function $u : X \to \mathbb{R}$ and a collection of probability distributions $\{\mu, \mu_f | f \in \mathcal{F}\}$ such that $\succsim$ ($\succsim_f$) has an expected utility representation. Additionally, $\mu(s) > 0$ for every $s \in S$, and for each f, $\mu_f(s) > 0$ for all $s \in A$.*

Proof. By Consistent expected utility it is standard to show the existence of (u, μ) and (u_f, μ_f) that represent $\succsim$ and $\succsim_f$, respectively. By ordinal preference consistency, we know that $u(x) \geq u(y)$ if and only if $u_f(x) \geq u_f(y)$, hence it follows that u_f is a positive affine transformation of u, so we simply apply the normalization that $u_f := u$. It also follows from monotonicity that $\mu(s) > 0$ for all s and $\mu_f(s) > 0$ for all $s \in S$. □

Let V, V_f denote the linear functionals generated by (u, μ), (u, μ_f), that represent $\succsim$ and $\succsim_f$, respectively. Note that these functionals are normalized by u so that $V(x) = V_f(x) = u(x)$.

Appendix B: Proofs of main results

B.1 *Proof of Theorem 1*

Proof of sufficiency. Assume that the collection of preferences $\{\succsim, \succsim_f\}$ satisfy Consistent expected utility, and Wishful Thinking. First, I will establish a key lemma.

Lemma 3. *If the collection of preferences $\{\succsim, \succsim_f\}$ satisfy Consistent expected utility, and Wishful Thinking, then preferences satisfy Similar state consistency.*

Proof. Fix any s, $\tilde{s}$ such that $f(s) \sim f(\tilde{s})$. Then fix x, y, z such that $x, y \succ z$ and $x\{s\}z \sim y\{\tilde{s}\}z$. Since $f(s) \succsim f(\tilde{s})$ and $x\{s\}z \succsim y\{\tilde{s}\}z$, then by Wishful Thinking it follows that $x\{s\}z \succsim_f y\{\tilde{s}\}z$. However, by symmetry it also follows from Wishful Thinking that $y\{\tilde{s}\}z \succsim_f x\{s\}z$, and hence $x\{s\}z \sim_f y\{\tilde{s}\}z$. From this, it follows that $u(x)\mu(s) + u(z)(1 - \mu(s)) = u(y)\mu(\tilde{s}) + u(z)(1 - \mu(\tilde{s}))$ and $u(x)\mu_f(s) + u(z)(1 - \mu_f(s)) = u(y)\mu_f(\tilde{s}) + u(z)(1 - \mu_f(\tilde{s}))$. After algebra, we conclude that $\frac{\mu(s)}{\mu(\tilde{s})} = \frac{\mu_f(s)}{\mu_f(\tilde{s})}$.

Now fix $B \subset S$ such that $f(s) \sim f(\tilde{s})$ for all $s, \tilde{s} \in B$. Since the above holds for all $s, \tilde{s} \in B$, we have $\mu_f(\tilde{s})\mu(s) = \mu_f(s)\mu(\tilde{s}) \Leftrightarrow \sum_{\tilde{s} \in B} \mu_f(\tilde{s})\mu(s) = \sum_{\tilde{s} \in B} \mu_f(s)\mu(\tilde{s}) \Leftrightarrow \mu_f(B)\mu(s) = \mu_f(s)\mu(B) \Leftrightarrow \frac{\mu_f(s)}{\mu_f(B)} = \frac{\mu(s)}{\mu(B)}$.

Now for any $h, g, z \in \mathcal{F}$, $hBz \succsim gBz$ is equivalent to

$$\begin{aligned}
\sum_{s\in B} u\big(h(s)\big)\mu(s) + \big(1-\mu(B)\big)u(z) \geq \sum_{s\in B} u\big(g(s)\big)\mu(s) + \big(1-\mu(B)\big)u(z) \quad &\Leftrightarrow \\
\sum_{s\in B} u\big(h(s)\big)\mu(s) \geq \sum_{s\in B} u\big(g(s)\big)\mu(s) \quad &\Leftrightarrow \\
\frac{1}{\mu(B)}\sum_{s\in B} u\big(h(s)\big)\mu(s) \geq \frac{1}{\mu(B)}\sum_{s\in B} u\big(g(s)\big)\mu(s) \quad &\Leftrightarrow \\
\frac{1}{\mu_f(B)}\sum_{s\in B} u\big(h(s)\big)\mu_f(s) \geq \frac{1}{\mu_f(B)}\sum_{s\in B} u\big(g(s)\big)\mu_f(s) \quad &\Leftrightarrow \\
\sum_{s\in B} u\big(h(s)\big)\mu_f(s) \geq \sum_{s\in B} u\big(g(s)\big)\mu_f(s).&
\end{aligned}$$

The last inequality is equivalent to $hBz \succsim_f gBz$. Since B was arbitrary, the result holds. ◁

By taking $B = S$, we have the immediate corollary.

Corollary 4. *For every constant act $x \in \mathcal{F}$, $\mu_x = \mu$.*

For all f which are nonconstant, define $\psi(f,s) := \frac{\mu_f(s)}{\mu(s)}$ for some x. By definition, it is clear that

$$\sum_{s\in S} u\big(g(s)\big)\psi(f,s)\mu(s) = \sum_{s\in S} u\big(g(s)\big)\left(\frac{\mu_f(s)}{\mu(s)}\right)\mu(s) = \sum_{s\in S} u\big(h(s)\big)\mu_f(s).$$

Suppose that $f(s) \sim f(\tilde{s})$. Then by Lemma 3, $\frac{\mu_f(s)}{\mu_f(\tilde{s})} = \frac{\mu(s)}{\mu(\tilde{s})}$, and thus

$$\psi(f,s) = \frac{\mu_f(s)}{\mu(s)} = \frac{\mu_f(\tilde{s})}{\mu(\tilde{s})} = \psi(f,\tilde{s})$$

For any f, a *payoff decomposition* of f is a partition of S, $\{E_1, \dots, E_n\}$, such that for all $s, \tilde{s} \in E_i$, $f(s) \sim f(\tilde{s})$ and for $i < j$ and any $s \in E_i$ and $\tilde{s} \in E_j$, $f(s) \succ f(\tilde{s})$. Then define $\psi_f(E_i) = \psi(f,s)$ for some $s \in E_i$. By the result above, this is well-defined. Next, E_k, E_{k+1} satisfy the conditions of Wishful Thinking. Thus let x, y, z satisfy $xE_kz \sim yE_{k+1}z$ with $x, y \succ z$. Since X is convex and u is only unique up to positive-affine transformation, we can assume $[0,1] \subset u(X)$, thus such x, y, z always exist. Thus $u(x)\mu(E_k) + u(z)(1-\mu(E_k)) = u(y)\mu(E_{k+1}) + u(z)(1-\mu(E_{k+1}))$, hence $(u(x)-u(z))\mu(E_k) = (u(y)-u(z))\mu(E_{k+1})$ and by Wishful Thinking it follows that $xE_kz \succsim_f yE_{k+1}z$. This is equivalent to

$$\begin{aligned}
&u(x)\mu(E_k)\psi_f(E_k) + u(z)\big[1-\mu(E_k)\psi_f(E_k)\big] \\
&\quad \geq u(y)\mu(E_{k+1})\psi_f(E_{k+1}) + u(z)\big[1-\mu(E_{k+1})\psi_f(E_{k+1})\big] \quad \Leftrightarrow \\
&\big(u(x)-u(z)\big)\mu(E_k)\psi_f(E_k) \geq \big(u(y)-u(z)\big)\mu(E_{k+1})\psi_f(E_{k+1}) \quad \Leftrightarrow
\end{aligned}$$

Or equivalently, $\psi_f(E_k) \geq \psi_f(E_{k+1})$. Next, define $\delta_f : u(X) \to (0, \infty)$ by $\delta_f(u(f(s))) = \psi(f, s)$ and define δ_f outside of $f(S)$ so that it is increasing. □

Proof of necessity. Suppose there exists μ, u and δ_f such that the representation holds, and define preferences in the usual way. Clearly, Consistent expected utility holds since the agent is a subjective expected utility maximizer. It remains to show the necessity of Wishful Thinking. Consider some f and suppose $xAz \succsim yBz$ for appropriate outcomes x, y, z, and events A, B. Then, since $xAz \succsim yBz$ it follows that $\mu(A)u(x) + (1-\mu(A))u(z) \geq \mu(B)u(y) + (1-\mu(B))u(z)$, or equivalently, $\frac{\mu(A)}{\mu(B)} \geq \frac{u(y)-u(z)}{u(x)-u(z)} \geq 0$. Thus it is sufficient to show that $\frac{\mu_f(A)}{\mu_f(B)} \geq \frac{\mu(A)}{\mu(B)}$. Define $\underline{d}_A = \min_{s\in A} \delta_f(u(f(s)))$ and $\overline{d}_B = \max_{s\in B} \delta_f(u(f(s)))$. Since A, B satisfy $\min_{s\in A} u(f(s)) \geq \max_{s\in B} u(f(s))$ and δ_f is increasing, it follows that $\underline{d}_A \geq \overline{d}_B$. Then it follows that

$$\mu_f(A) = \sum_{s\in A} \delta_f(u\big(f(s)\big)\mu(s) \geq \underline{d}_A \sum_{s\in A} \mu(s)$$

and

$$\overline{d}_B \sum_{s\in B} \mu(s) \geq \sum_{s\in B} \delta_f(u\big(f(s)\big)\mu(s) = \mu_f(B).$$

Hence

$$\frac{\mu_f(A)}{\mu_f(B)} \geq \frac{\underline{d}_A \sum_{s\in B} \mu(s)}{\overline{d}_B \sum_{s\in B} \mu(s)} \geq \frac{\mu(A)}{\mu(B)},$$

and thus $xBz \succsim_f yCz$, completing the proof. □

B.2 *Proof of Theorem 2*

Proof of sufficiency. Assume Independence of irrelevant payoffs. The proof proceeds by constructing a value function v such that δ_f is a Consequential distortion relative to v.

Lemma 4. *For all $f, g \in \mathcal{F}$ and all $s, \tilde{s} \in S$ if $f(s) \sim g(s)$ and $f(\tilde{s}) \sim g(\tilde{s})$, then $\frac{\psi(f,s)}{\psi(f,\tilde{s})} = \frac{\psi(g,s)}{\psi(g,\tilde{s})}$.*

Proof. Let $A = \{s, \tilde{s}\}$ and suppose $f(s) \sim g(s)$ and $f(\tilde{s}) \sim g(\tilde{s})$. Then for all h, j, z, $hAz \succsim_f jAz$ if and only if

$$\begin{aligned} &u\big(h(s)\big)\mu_f(s) + u\big(h(\tilde{s})\big)\mu_f(\tilde{s}) + u(z)\big(1-\mu_f(A)\big) \\ &\quad \geq u\big(j(s)\big)\mu_f(s) + u\big(j(\tilde{s})\big)\mu_f(\tilde{s}) + u(z)\big(1-\mu_f(A)\big) \quad \Leftrightarrow \\ &\quad \big[u\big(h(s)\big) - u\big(j(s)\big)\big]\mu_f(s) \geq \big[u\big(j(\tilde{s})\big) - u\big(h(\tilde{s})\big)\big]\mu_f(\tilde{s}) \end{aligned}$$

Suppose h, j are such that $hAz \sim_f jAz$. Then by Independence of irrelevant payoffs, $hAz \sim_g jCz$, and hence

$$\frac{\mu_f(s)}{\mu_f(\tilde{s})} = \frac{u\big(j(\tilde{s})\big) - u\big(h(\tilde{s})\big)}{u\big(h(s)\big) - u\big(j(s)\big)} = \frac{\mu_g(s)}{\mu_g(\tilde{s})}.$$

Since $\psi(f,s) := \frac{\mu_f(s)}{\mu(s)}$, it follows that

$$\frac{\psi(f,s)}{\psi(f,\tilde{s})} = \frac{\mu_f(s)}{\mu_f(\tilde{s})} \times \frac{\mu(\tilde{s})}{\mu(s)} = \frac{\mu_g(s)}{\mu_g(\tilde{s})} \times \frac{\mu(\tilde{s})}{\mu(s)} = \frac{\psi(g,s)}{\psi(g,\tilde{s})}. \qquad \triangleleft$$

Define the function $\phi : X \times X \to \mathbb{R}_+$ by $\phi(x,y) := \frac{\psi(f,s)}{\psi(f,\tilde{s})}$ for some f with $f(s) = x$ and $f(\tilde{s}) = y$. By the previous lemma, for all f and g such that $f(s) = x = g(s)$ and $f(\tilde{s}) = y = g(\tilde{s})$, $\frac{\psi(f,s)}{\psi(f,\tilde{s})} = \frac{\psi(g,s)}{\psi(g,\tilde{s})}$, hence ϕ is well-defined.

Lemma 5. *The function ϕ satisfies the following properties: (i) $x \succsim y \implies \phi(x,y) \geq 1$, (ii) $\phi(x,y)\phi(y,z) = \phi(x,z)$, (iii) $\frac{1}{\phi(x,y)} = \phi(y,x)$, and (iv) $\phi(x,x) = 1$.*

Proof. *(i)* Fix s, $\tilde{s}$ such that $f(s) = x \succsim y = f(\tilde{s})$. By the previous theorem $\psi(f,s) \geq \psi(f,\tilde{s})$, hence $\phi(x,y) = \frac{\psi(f,s)}{\psi(f,\tilde{s})} \geq 1$. *(ii)* Fix three states s_x, s_y, s_z, where $f(s_i) = i$, $i \in \{x,y,z\}$. Then $\phi(x,y)\phi(y,z) = \frac{\psi(f,s_x)}{\psi(f,s_y)}\frac{\psi(f,s_y)}{\psi(f,s_z)} = \phi(x,z)$. *(iii)* For any s, $\tilde{s}$ with $f(s) = x$, $f(\tilde{s}) = y$, $\frac{1}{\phi(x,y)} = \frac{1}{\frac{\psi(f,s)}{\psi(f,\tilde{s})}} = \frac{\psi(f,\tilde{s})}{\psi(f,s)} = \phi(y,x)$. *(iv)* It follows directly from *(iii)* that $\phi(x,x) = \frac{1}{\phi(x,x)}$, hence $\phi(x,x)\phi(x,x) = \phi(x,x) = 1$. $\triangleleft$

Fix some $x_* \in X$ and define $v : u(X) \to \mathbb{R}_+$ by $v(a) := \phi(u^{-1}(a), x_*)$.

Then for any f such that $x = f(s)$ and $y = f(\tilde{s})$ for some s, $\tilde{s}$,

$$\frac{v\big(u(x)\big)}{v\big(u(y)\big)} = \frac{\phi(x,x_*)}{\phi(y,x_*)} = \phi(x,y) = \frac{\psi(f,s)}{\psi(f,\tilde{s})} = \frac{\mu_f(s)}{\mu_f(\tilde{s})}\frac{\mu(\tilde{s})}{\mu(s)} \quad \Leftrightarrow$$

$$\frac{v\big(u(x)\big)}{v\big(u(y)\big)}\frac{\mu(s)}{\mu(\tilde{s})} = \frac{\mu_f(s)}{\mu_f(\tilde{s})} \quad \Leftrightarrow$$

$$\mu_f(\tilde{s}) = \frac{\mu_f(s)}{v\big(u\big(f(s)\big)\big)\mu(s)} v\big(u\big(f(\tilde{s})\big)\big)\mu(\tilde{s})$$

Summing over $\tilde{s} \in S$ yields

$$1 = \sum_{\tilde{s}} \mu_f(\tilde{s}) = \left(\sum_{\tilde{s}} v\big(u\big(f(\tilde{s})\big)\big)\mu(\tilde{s})\right)\frac{\mu_f(s)}{v\big(u\big(f(s)\big)\big)\mu(s)}$$

hence

$$\mu_f(s) = \frac{v\big(u\big(f(s)\big)\big)}{\sum_{\tilde{s}} v\big(u\big(f(\tilde{s})\big)\big)\mu(\tilde{s})}\mu(s).$$

Lemma 6. *v is increasing.*

Proof. Suppose $a, b \in u(X)$ and $a \geq b$. Then for some $x \succsim y$, $u(x) = a$, $u(y) = b$. Then $\frac{v(a)}{v(b)} = \frac{\phi(u^{-1}(a), x_*)}{\phi(u^{-1}(b), x_*)} = \phi(x, x_*)\phi(x_*, y) = \phi(x, y) \geq 1$, hence $v(a) \geq v(b)$. ◁

By combining the preceding lemmas, the result follows. □

Proof of Necessity. Since the Consequential distortion is a special case of the Wishful Thinking Representation, we must only prove the necessity of Independence of irrelevant payoffs. Assume the Consequential distortion holds for some v. Then consider any two scenarios f and g, and some $A \subset S$ such that $f(s) \sim g(s)$ for all $s \in A$. Then it follows that $u(f(s)) = u(g(s))$ for all $s \in A$, and thus for any two $s, \tilde{s} \in A$,

$$\frac{\mu_f(s)}{\mu_f(\tilde{s})} = \frac{v\big(u\big(f(s)\big)\big)}{v\big(u\big(f(\tilde{s})\big)\big)} \frac{\mu(s)}{\mu(\tilde{s})} = \frac{v\big(u\big(g(s)\big)\big)}{v\big(u\big(g(\tilde{s})\big)\big)} \frac{\mu(s)}{\mu(\tilde{s})} = \frac{\mu_g(s)}{\mu_g(\tilde{s})}.$$

Through basic algebra and then summing over $\tilde{s} \in A$, we conclude that $\frac{\mu_f(s)}{\mu_f(A)} = \frac{\mu_g(s)}{\mu_g(A)}$ for every $s \in A$. If follows then that for any $h, j \in \mathcal{F}$ and $z \in X$, $hCz \succsim_f jCz$ if and only if $\sum_{s\in A} \mu_f(s)u(h(s)) + (1 - \mu_f(A))u(z) \geq \sum_{s\in A} \mu_f(s)u(j(s)) + (1 - \mu_f(A))u(z)$. Algebra yields $\sum_{s\in A} \frac{\mu_f(s)}{\mu_f(A)} u(h(s)) \geq \sum_{s\in A} \frac{\mu_f(s)}{\mu_f(A)} u(j(s))$. Now, substituting $\frac{\mu_f(s)}{\mu_f(A)} = \frac{\mu_g(s)}{\mu_g(A)}$ provides $\sum_{s\in A} \frac{\mu_g(s)}{\mu_g(A)} u(h(s)) \geq \sum_{s\in A} \frac{\mu_g(s)}{\mu_g(A)} u(j(s))$. From here, it is straightforward to conclude $hCz \succsim_g jCz$, and thus Independence of irrelevant payoffs holds. □

B.3 *Proof of Theorem 3*

It is straightforward to see that Wishful Thinking implies both Similar state consistency (see Lemma 3) and Minimal status quo bias, so I will only show that (ii) implies (i). Consider any f and suppose events A and B and outcomes x, y, z satisfy the conditions of Wishful Thinking and suppose that $xAz \succsim yBz$. Let $\{E_1, \dots, E_n\}$ be a decomposition of f and suppose $x_i = f(s)$ for some $s \in E_i$. For some $i < j$, let

$$g(s) = \begin{cases} x_i & \text{if } s \in E_i \\ x_j & \text{if } s \in E_j \\ z & \text{if } s \in S \setminus (E_i \cup E_j) \end{cases}$$

where z satisfies $x_i \succ z \succ x_j$, and $z \sim f\{E_i \cup E_j\}z$. By Minimal status quo bias, if follows that $g \succsim_g z$, and hence $\frac{\mu_g(E_i)}{\mu_g(E_j)} \geq \frac{\mu(E_i)}{\mu(E_j)}$, while by Independence of irrelevant payoffs it follows that $\frac{\mu_g(E_i)}{\mu_g(E_j)} = \frac{\mu_f(E_i)}{\mu_f(E_j)}$, the proof of which is similar to Case 1 of Lemma 4. Combining yields that $\frac{\mu_f(E_i)}{\mu_f(E_j)} \geq \frac{\mu(E_i)}{\mu(E_j)}$ whenever $i < j$. Thus there exists numbers $\lambda(f, i)$ such that

$\mu_f(E_i) = \lambda(f,i)\mu(E_i)$, where $\lambda(f,i) \geq \lambda(f,j)$ when $i < j$. Finally,

$$\frac{\mu_f(A)}{\mu_f(B)} = \frac{\sum_{E_i : E_i \cap A \neq \varnothing} \lambda(f,i)\mu(A \cap E_i)}{\sum_{E_j : E_j \cap B \neq \varnothing} \lambda(f,j)\mu(B \cap E_j)} \geq \frac{\mu(A)}{\mu(B)},$$

where the last inequality follows from the fact that $\min_{i:E_i \cap A \neq \varnothing} \lambda(f,i) \geq \max_{j:E_j \cap B \neq \varnothing} \lambda(f, j)$. To complete the argument, observe that $xAz \succsim yBz$ is equivalent to $\frac{\mu(A)}{\mu(B)} \geq \frac{u(y)-u(z)}{u(y)-u(z)}$. Since $\frac{\mu_f(A)}{\mu_f(B)} \geq \frac{\mu(A)}{\mu(B)}$, it follows that $xAz \succsim_f yBz$, and Wishful Thinking holds.

B.4 *Proof of Theorem 4*

Proof of sufficiency. Suppose Best-case dominance holds. First, note that for all constant acts, $f \in \mathcal{C} \equiv \{f \in \mathcal{F} \mid \text{for some } x \in X, f(s) = x \text{ for all } s \in S\}$, $\mathcal{D}(f) = S$, hence for any $\delta \in [0,1]$, $\mu_f = \mu = (1-\delta)\mu + \delta\mu_{|\mathcal{D}(f)}$. For the remainder of the proof, suppose f is nonconstant: $f \in \mathcal{N} \equiv \mathcal{F} \setminus \mathcal{C}$. Let $H(f) := C(f) \cup X$, where $C(f) = \{h | h \asymp f\}$ and X is understood to mean the set of constant acts. Note that $H(f) \subset \mathrm{cl}(C(f))$. The proof will be shown via the following lemmas.

Lemma 7. *There exists a function $\delta : \mathcal{N} \to [0,1]$ such that for all $h, g \in \mathrm{cl}(C(f))$, $h \succsim_f g$ if and only if $(1-\delta(f))V(h) + \delta(f))V_{\mathcal{D}(f)}(h) \geq (1-\delta(f))V(g) + \delta(f)V_{\mathcal{D}(f)}(g)$.*

Proof. Fix some $f \in \mathcal{N}$.

Next, define $\trianglerighteq$ by $h \trianglerighteq g$ if and only if for some $s \in S$, $h(s) \succsim g(s')$ for all $s' \in S$. Let $\triangleright$ and $\approx$ denote the strict and symmetric parts of $\trianglerighteq$. Further, $\trianglerighteq$ is represented by $M(h) = \max\{u(h(s)) | s \in S\}$. Suppose $h \trianglerighteq g$. Then for some $\hat{s} \in S$, $h(\hat{s}) \succsim g(s')$ for all $s' \in S$. Hence $\max\{u(h(s)) | s \in S\} \geq u(h(\hat{s})) \geq \max\{u(g(s)) | s \in S\}$. Next, suppose $\max\{u(h(s)) | s \in S\} \geq \max\{u(g(s)) | s \in S\}$. Then let s^* solve $u(h(s^*)) = \max\{u(h(s)) | s \in S\}$. Then clearly $h(s^*) \succsim g(s)$ for all $s \in S$.

If $h, g \in C(f)$, it follows that $\mathcal{D}(h) = \mathcal{D}(g) = \mathcal{D}(f)$, hence $h \trianglerighteq g$ is equivalent to $h(s) \succsim g(s)$ for all $s \in \mathcal{D}(f)$. By the previous lemma, for $h, g \in C(f)$ there is some x_h, x_g such that $h(s) \sim x_h$ and $g(s) \sim x_g$ for all $s \in \mathcal{D}(f)$. It is then clear that $h \trianglerighteq g \Leftrightarrow x_h \succsim x_g$ for $h, g \in C(f)$, or equivalently, $u(x_h) = \max\{u(h(s)) | s \in S\}$. Then for any $\rho \in \Delta(S)$ satisfying $\rho(\mathcal{D}(f)) = 1$, it follows that $\sum_{s \in \mathcal{D}(f)} u(h(s))\rho(s) = u(x_h) = M(h)$. Further, for any $x \in \mathcal{F}$, $\sum_{s \in \mathcal{D}(f)} u(x(s))\rho(s) = u(x) = M(x)$, and hence

$$U_\rho(h) := \sum_{s \in \mathcal{D}(f)} u\big(h(s)\big)\rho(s)$$

represents $\trianglerighteq$ on $H(f)$, and U_ρ is a normalized linear functional on $\mathcal{F}$.

For $\rho \in \Delta(S)$ such that $\rho(\mathcal{D}(f)) = 1$, define the set of utility values $\mathcal{U}_\rho := \{(U_\rho(h), V(h)) \in \mathbb{R}^2 | h \in H(f)\}$. Then for each pair of utility values, $(v_1, v_2), (v_1', v_2') \in \mathcal{U}$, define $\succsim^*$ by

$$(v_1, v_2) \succsim_f^* \big(v_1', v_2'\big) \quad \Leftrightarrow \quad h \succsim_f g$$

for some $h, g \in H(f)$ such that $(U_\rho(h), V(h)) = (v_1, v_2)$ and $(U_\rho(g), V(g)) = (v_1', v_2')$. This relation is well-defined, since $(U_\rho(h), V(h)) = (U_\rho(g), V(g))$ implies $h \approx g$ and $h \cong g$, hence $h \sim_f g$. Further, for all $h \in H(f)$, $U_\rho(h) \geq V(h)$, hence $v_1 \geq v_2$ for all $(v_1, v_2) \in \mathcal{U}$. This holds because U_ρ coincides with the maximal payoff of h in S. Further, it is obvious that $\succsim^*$ is complete, transitive, monotonic, and satisfies independence and continuity. Let $\bar{s} \in \mathcal{D}(f)$ and $\underline{s} \in \arg\min\{u(f(s)) | s \in S\}$. Then since $f \in \mathcal{N}$, the constant acts $f(\bar{s})$ and $f(\underline{s})$ satisfy $f(\bar{s}) \succ f(\underline{s})$. Further, for $\alpha \in (0, 1)$, $h := \alpha f + (1 - \alpha) f(\bar{s})$ satisfies $u(f(\bar{s})) > U_\rho(h) > V(h) > u(f(\underline{s}))$, hence there are $(v_1^*, v_2^*), (\bar{v}, \bar{v}), (\underline{v}, \underline{v}) \in \mathcal{U}$ such that $v_1^* > v_2^*$ and $\bar{v} > v_1^* > \underline{v}$, where this follows due to the convexity of $H(f)$ and the fact that U_ρ and V are normalized. Hence by Lemma 2 of Saito (2013), there exists some $\delta_f \in [0, 1]$ such that $\delta_f U_\rho(h) + (1 - \delta_f) V(h) \geq \delta_f U_\rho(g) + (1 - \delta_f) V(g) \Leftrightarrow$

$$\delta v_1 + (1 - \delta) v_2 \geq \delta v_1' + (1 - \delta) v_2' \quad \Leftrightarrow \quad (v_1, v_2) \succsim^* \left(v_1', v_2'\right) \quad \Leftrightarrow \quad h \succsim_f g.$$

By Lemma 2, for every ρ, $W_\rho := (1 - \delta) V(h) + \delta U_\rho(h)$ is a normalized linear functional given by $\mu_W = (1 - \delta_f)\mu_{|A} + \delta_f \rho$. Then by continuity of W_ρ we extend it to $\mathrm{cl}(C(f))$. If $\mathcal{D}(f)$ is a singleton then ρ is uniquely given. Suppose then without loss that $|\mathcal{D}(f)| \geq 2$ and fix $s, s' \in \mathcal{D}(f)$. Consider x, y, z such that $x, y \succ z$ and $x\{s\}z \sim y\{s'\}z$. Then define $h := x\{s\}z$, $g := y\{s'\}z$ and $B := \{s, s'\}$. Then by $hBz \sim gBz$, and by Lemma 3, $hBz \sim_f gBz \Leftrightarrow x\{s\}z \sim_f y\{s'\}z$. W_ρ satisfies the equation $W_\rho(x\{s\}z) = W_\rho(y\{s'\}z)$ if and only if $\mu_W(s)u(x) + (1 - \mu_W(s))u(z) = \mu_W(s')u(y) + (1 - \mu_W(s'))u(z)$. This is equivalent to $[u(x) - u(z)]\mu_W(s) = [u(y) - u(z)]\mu_W(s')$, and hence $\frac{\mu_W(s)}{\mu_W(s')} = \frac{u(y)-u(z)}{u(x)-u(z)}$. However, from $x\{s\}z \sim y\{s'\}z$ we also know that $\frac{\mu(s)}{\mu(s')} = \frac{u(y)-u(z)}{u(x)-u(z)}$, hence

$$\frac{(1 - \delta_f)\mu(s) + \delta_f \rho(s)}{(1 - \delta_f)\mu(s') + \delta_f \rho(s')} = \frac{\mu_W(s)}{\mu_W(s')} = \frac{\mu(s)}{\mu(s')}.$$

Algebra yields $\frac{\rho(s)}{\rho(s')} = \frac{\mu(s)}{\mu(s')}$, which when combined with $\rho(\mathcal{D}(f)) = 1$, implies $\rho = \mu_{|\mathcal{D}(f)}$ is the unique ρ such that W_ρ represents $\succsim_f$ on $\mathrm{cl}(C_A(f))$, hence $W_{\mu_{\mathcal{D}(f)}} = (1 - \delta_f) V(h) + \delta_f V_{\mathcal{D}(f)}(h)$. Since for each f the number δ_f is unique, we simply define the function $\delta : \mathcal{N} \to [0, 1]$ by $\delta(f) = \delta_f$. $\triangleleft$

Lemma 8. *For all $h, g \in \mathcal{F}$, $h \succsim_f g$ if and only if*

$$\big(1 - \delta(f)\big) V(h) + \delta(f) V_{\mathcal{D}(f)}(h) \geq \big(1 - \delta(f)\big) V_A(g) + \delta(f) V_{\mathcal{D}(f)}(g).$$

Proof. Again by Lemma 2, linearity of $(1 - \delta(f))V + \delta(f) V_{\mathcal{D}(f)}$ implies we can extend it to all of $\mathcal{F}$ by the equation

$$\big(1 - \delta(f)\big) V(h) + \delta(f) V_{\mathcal{D}(f)}(h) = \sum_{s \in S} u\big(h(s)\big)\big[\big(1 - \delta(f)\big)\mu + \delta(f)\mu_{|\mathcal{D}(f)}\big](s).$$

That is, we can define a linear functional on $\mathcal{F}$ by

$$U_f(h) = \sum_{s \in S} u\big(h(s)\big)\big[\big(1 - \delta(f)\big)\mu + \delta(f)\mu_{|\mathcal{D}(f)}\big](s).$$

Then U_f and V_f are both normalized linear functionals that agree on $\mathrm{cl}(C(f))$, and hence by uniqueness of subjective probabilities, it follows that $\mu_f = (1-\delta(f))\mu + \delta(f)\mu_{|\mathcal{D}(f)}$.

Consider any two states $s, s' \in S$. Then without loss $f(s) \succ f(s')$ or $f(s) \sim f(s')$. Suppose the first case holds, and for convenience, ignore the dependence of δ on f. Then for any $x \succ y$, $x\{s\}y \in \mathrm{cl}(C(f))$. Say for some $w \in X$, $x\{s\}y \sim_f w$, then it follows that $U_f(x\{s\}y) = u(w) = V_f(x\{s\}y)$, hence

$$\begin{aligned} &u(x)\mu_f(s) + u(y)\big(1-\mu_f(s)\big) \\ &\quad = u(x)\big[(1-\delta)\mu_A + \delta\mu_{|\mathcal{D}(f)}\big](s) + u(y)\big(1-\big[(1-\delta)\mu + \delta\mu_{|\mathcal{D}(f)}\big](s)\big). \end{aligned}$$

Since $u(x) > u(y)$, it immediately follows that $\mu_f(s) = [(1-\delta)\mu + \delta\mu_{|\mathcal{D}(f)}](s)$.

Next, it is immediately apparent that for $f(s) \sim f(s')$, $\frac{\mu_f(s)}{\mu_f(s')} = \frac{[(1-\delta)\mu+\delta\mu_{|\mathcal{D}(f)}](s)}{[(1-\delta)\mu+\delta\mu_{|\mathcal{D}(f)}](s')}$ since either $\{s, s'\} \subseteq \mathcal{D}(f)$ or $\{s, s'\} \nsubseteq \mathcal{D}(f)$. Hence in any case, for any $s, s' \in S$

$$\begin{aligned} \frac{\mu_f(s)}{\mu_f(s')} &= \frac{\big[(1-\delta)\mu + \delta\mu_{|\mathcal{D}(f)}\big](s)}{\big[(1-\delta)\mu + \delta\mu_{|\mathcal{D}(f)}\big](s')} \quad \Rightarrow \\ \mu_f(s)\big[(1-\delta)\mu + \delta\mu_{|\mathcal{D}(f)}\big](s') &= \mu_f(s')\big[(1-\delta)\mu + \delta\mu_{|\mathcal{D}(f)}\big](s) \quad \Rightarrow \\ \sum_{s'\in S}\mu_f(s)\big[(1-\delta)\mu + \delta\mu_{|\mathcal{D}(f)}\big](s') &= \sum_{s'\in S}\mu_f(s')\big[(1-\delta)\mu + \delta\mu_{|\mathcal{D}(f)}\big](s) \quad \Rightarrow \\ \mu_f(s) &= \big[(1-\delta)\mu + \delta\mu_{|\mathcal{D}(f)}\big](s). \end{aligned}$$

◁

The previous two lemmas prove sufficiency. □

Proof of necessity. Since the Best-Case Binary representation is a special case of the Wishful Thinking Representation, all that must be shown is the necessity of Best-case dominance. Suppose the Best-Case Binary representation holds. Pick any scenario f and any $h, g \asymp f$ such that $h \succsim g$ and $h(s) \succsim g(s')$ for some $s \in S$ and all $s' \in S$. It follows from previous results that $\mathcal{D}(f) = \mathcal{D}(h) = \mathcal{D}(g)$ and that for every $s \in \mathcal{D}(f)$, $u(h(s)) = \max_{s'\in S} u(h(s')) \geq \max_{s'\in S} u(g(s')) = u(g(s))$, and hence

$$\sum_{s\in S} u\big(h(s)\big)\mu\big(s|\mathcal{D}(f)\big) \geq \sum_{s\in S} u\big(g(s)\big)\mu\big(s|\mathcal{D}(f)\big). \tag{1}$$

Similarly, it follows from previous results that

$$\sum_{s\in S} u\big(h(s)\big)\mu(s) \geq \sum_{s\in S} u\big(g(s)\big)\mu(s). \tag{2}$$

Then by multiplying both sides of (1) by $\delta(f)$, multiplying both sides of (2) by $1-\delta(f)$, and adding we can conclude that

$$\begin{aligned} &\big(1-\delta(f)\big)\sum_{s\in S} u\big(h(s)\big)\mu(s) + \delta(f)\sum_{s\in S} u\big(h(s)\big)\mu\big(s|\mathcal{D}(f)\big) \\ &\quad \geq \big(1-\delta(f)\big)\sum_{s\in S} u\big(g(s)\big)\mu(s) + \delta(f)\sum_{s\in S} u\big(g(s)\big)\mu\big(s|\mathcal{D}(f)\big). \end{aligned}$$

This is equivalent to $V_f(h) \geq V_f(g)$, and hence $h \succsim_f g$. □

B.5 *Proof of Theorem 6*

Proof of Case 1: (Consequential distortion). ($\Rightarrow$)Let $x \succ y$ and suppose for some $A \subsetneq S$ and let $f = xAy$. Then for any z satisfying $f \succsim_f^1 z$, it follows that if $\succsim_f^2$ is more wishful than $\succsim_f^1$, then $V_f^2(xAy) \geq V_f^1(xAy)$. Hence for $B = S \backslash A$, it follows that $\frac{\mu_f^2(A)}{\mu_f^2(B)} \geq \frac{\mu_f^1(A)}{\mu_f^1(B)}$ and, therefore, $\frac{v^2(u(x))}{v^2(u(y))} \geq \frac{v^1(u(x))}{v^1(u(y))}$.

($\Leftarrow$) Suppose for all $a, b \in R$ with $a \geq b$, $\frac{v^2(a)}{v^2(b)} \geq \frac{v^1(a)}{v^1(b)}$. Clearly, the result is trivial if f is constant, so suppose that it is nonconstant and consider the decomposition of f: $\{E_1, \dots, E_n\}$. I will now construct an induced probability distribution over utilities, $\mathcal{U} \equiv \{u(f(s)) | s \in S\}$. Note that for each i, $u(f(s)) = a_i$ for all $s \in E_i$ and $a_i > a_j$ for $j < i$, by the properties of a decomposition. Define, for each person $k = 1, 2$, the distribution over $\mathcal{U}$ by $\pi^k(a_i) = \mu_f^k(E_i)$. Then whenever $a_i > a_j$, the distributions satisfy

$$\frac{\pi^2(a_i)}{\pi^2(a_j)} = \frac{\mu_f^2(E_i)}{\mu_f^2(E_i)} = \frac{v^2(a_i)}{v^2(a_j)} \frac{\mu(E_i)}{\mu(E_j)} \geq \frac{v^1(a_i)}{v^1(a_j)} \frac{\mu(E_i)}{\mu(E_j)} = \frac{\mu_f^1(E_i)}{\mu_f^1(E_i)} = \frac{\pi^1(a_i)}{\pi^1(a_j)}.$$

Hence π^2 and π^1 have monotone likelihood ratios and hence π^2 first-order stochastically dominates π^1. But this implies that

$$V_f^2(f) = \sum_{s \in S} u\big(f(s)\big)\mu_f^2(s) = \sum_i a_i \pi^2(a_i) \geq \sum_i a_i \pi^1(a_i) = \sum_{s \in S} u\big(f(s)\big)\mu_f^1(s) = V_f^1(f).$$

From this, it clearly follows that $f \succsim_f^1 x \Rightarrow f \succsim_f^2 x$. □

Proof of Case 2: (Best-case binary distortion). ($\Rightarrow$) Since $\succsim^1 = \succsim^2$, it follows that $u^1 = u^2 = u$ and $\mu^1 = \mu^2 = \mu$. Suppose for all f, $f \succsim_f^1 x \Rightarrow f \succsim_f^2 x$, but $\delta^2 < \delta^1$. Let $f \in \mathcal{N}$ and pick $\bar{x} \sim_f^1 f$. Hence $(1 - \delta^1) \sum_{s \in S} u(f(s))\mu(s) + \delta^1 \sum_{s \in S} u(f(s))\mu(s|\mathcal{D}(f)) = u(\bar{x})$. Since $\delta^2 < \delta^1$ and $\sum_{s \in S} u(f(s))\mu(s|\mathcal{D}(f)) > \sum_{s \in S} u(f(s))\mu(s)$, it follows that $u(\bar{x}) > (1 - \delta^2) \sum_{s \in S} u(f(s))\mu(s) + \delta^2 \sum_{s \in S} u(f(s))\mu(s|\mathcal{D}(f))$. But this contradicts $f \succsim_f^2 \bar{x}$, thus $\delta^2(f) \geq \delta^1(f)$ for all f.

($\Leftarrow$) Suppose $\delta^2(f) \geq \delta^1(f)$ for all f, and without loss suppose $f \in \mathcal{N}$. Then since $\sum_{s \in S} u(f(s))\mu(s|\mathcal{D}(f)) > \sum_{s \in A} u(f(s))\mu(s)$, it clearly follows from the representation that $V_f^2(f) \geq V_f^1(f)$, and hence $f \succsim_f^1 x \Rightarrow f \succsim_f^2 x$. □

Appendix C: Applications

Proof of Proposition 1. It is simple to show that utility maximizing demands are $a_i^* = \mu_{\omega_i}(A)\frac{w_i}{p_a}$ and $b_i^* = \mu_{\omega_i}(B)\frac{w_i}{p_b}$, respectively. Hence the market excess demand for a is given by $z(a) = \sum_i \mu_{\omega_i}(A)\frac{w_i}{p_a} - \sum_i \omega_{ia} = \sum_i \mu_{\omega_i}(A)\frac{w_i}{p_a} - 1$. Since endowments are interior and beliefs are fixed by the endowment, one can simply normalize prices so that $p_b^* = 1 - p_a^*$ and use the market clearing condition to solve for prices. As endowments are completely symmetric, $\omega_{1a} = \omega_{2b}$, $\omega_{ia} + \omega_{ib} = 1$ for each i. Further,

$\delta_{\omega_1}(x) = \delta_{\omega_2}(x) \equiv \delta(x)$ since agent's have identical preferences conditional on endowments and endowments are symmetric. Hence, simple algebra results in $p_a^* = \frac{1}{2}$ and $w_i = \frac{1}{2}$ for each i. Thus $(a_i^*, b_i^*) = (\frac{\delta(\omega_{ia})}{\delta(\omega_{ia})+\delta(\omega_{ib})}, \frac{\delta(\omega_{ib})}{\delta(\omega_{ib})+\delta(\omega_{ia})})$.

Then by using the functional form of δ, we conclude $(a_i^*, b_i^*) = (\frac{\omega_{ia}^\alpha}{\omega_{ia}^\alpha+\omega_{ib}^\alpha}, \frac{\omega_{ib}^\alpha}{\omega_{ia}^\alpha+\omega_{ib}^\alpha})$. For $\alpha = 0$, the result is obvious. For $0 < \alpha < 1$ and, without loss, $\omega_{1a} > \frac{1}{2}$, it follows that $\frac{1}{2} < a_1^* < \omega_{1a}$ and $\frac{1}{2} > b_1^* > \omega_{1b}$, and hence there is partial risk sharing. For $\alpha = 1$, it follows that $(a_i^*, b_i^*) = (\omega_{ia}, \omega_{ib})$ and there is no trade. For $\alpha > 1$ and, without loss, $\omega_{1a} > \frac{1}{2}$, $a_1^* > \omega_{1a}$ and $b_1^* < \omega_{1b}$, and hence agents take additional risk. □

Proof of Proposition 2. This follows immediately from Proposition 1 by substituting $\mu_{\omega_1}(A) = \frac{\omega_{1a}^\alpha}{\omega_{1a}^\alpha+\omega_{1b}^\alpha}$, $\mu_{\omega_1}(B) = 1 - \mu_{\omega_1}(A)$, and $\mu_{\omega_2}(A) = \mu_{\omega_2}(B) = \frac{1}{2}$ into the formula for excess demand and solving. □

References

Anscombe, Francis J. and Robert J. Aumann (1963), "A definition of subjective probability." *Annals of Mathematical Statistics*, 34, 199–205. [993]

Apesteguia, Jose and Miguel A. Ballester (2009), "A theory of reference-dependent behavior." *Economic Theory*, 40, 427–455. [1009]

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer (1995), "Biased judgments of fairness in bargaining." *American Economic Review*, 85, 1337–1343. [992]

Bénabou, Roland (2013), "Groupthink: Collective delusions in organizations and markets." *Review of Economic Studies*, 80, 429–462. [1009]

Bénabou, Roland and Jean Tirole (2011), "Identity, morals, and taboos: Beliefs as assets." *Quarterly Journal of Economics*, 126, 805–855. [1009]

Bewley, Truman F. (2002), "Knightian decision theory: Part I." *Decisions in Economics and Finance*, 25, 79–110. [1009]

Brunnermeier, Markus K. and Jonathan A. Parker (2005), "Optimal expectations." *American Economic Review*, 95, 1092–1118. [1008]

Budescu, David and Meira Bruderman (1995), "The relationship between the illusion of control and the desirability bias." *Journal of Behavioral Decision Making*, 8, 109–125. [992]

Caplin, Andrew and John Leahy (2001), "Psychological expected utility theory and anticipatory feelings." *Quarterly Journal of Economics*, 116, 55–79. [1009]

Cohen, Lauren (2009), "Loyalty-based portfolio choice." *Review of Financial Studies*, 22, 1213–1245. [992, 994]

Dean, Mark, Özgür Kibris, and Yusufcan Masatlioglu (2017), "Limited attention and status quo bias." *Journal of Economic Theory*, 169, 93–127. [990, 996, 1009]

DiTella, Rafael, Sebastian Galiant, and Ernesto Schargrodsky (2007), "The formation of beliefs: Evidence from the allocation of land titles to squatters." *Quarterly Journal of Economics*, 122, 209–241. [992]

Eliaz, Kfir and Ran Spiegler (2011), "On the strategic use of attention grabbers." *Theoretical Economics*, 6, 127–155. [990]

Epstein, Larry G. and Igor Kopylov (2007), "Cold feet." *Theoretical Economics*, 2, 231–259. [1009]

Eyster, Erik (2002), "Rationalizing the past: A taste for consistency." Unpublished paper, London School of Economics. [1009]

Isoni, Andrea, Graham Loomes, and Robert Sugden (2011), "The willingness to pay—willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations: Comment." *American Economic Review*, 101, 991–1011. [1004]

Kahneman, Daniel and Amos Tversky (1979), "Prospect theory: An analysis of decision under risk." *Econometrica*, 47, 263–291. [990]

Knetsch, Jack L. (1989), "The endowment effect and evidence of nonreversible indifference curves." *American Economic Review*, 79, 1277–1284. [1004]

Kőszegi, Botond and Matthew Rabin (2006), "A model of reference-dependent preferences." *Quarterly Journal of Economics*, 121, 1133–1165. [1008]

Kunda, Ziva (1987), "Motivated inference: Self-serving generation and evaluation of causal theories." *Journal of Personality and Social Psychology*, 53, 636–647. [992]

Langer, Ellen J. (1975), "The illusion of control." *Journal of Personality and Social Psychology*, 32, 311–328. [992]

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006), "Ambiguity aversion, robustness, and the variational representation of preferences." *Econometrica*, 74, 1447–1498. [993]

Masatlioglu, Yusufcan and Efe A. Ok (2005), "Rational choice with status quo bias." *Journal of Economic Theory*, 121, 1–29. [990, 991, 993, 1005, 1009]

Masatlioglu, Yusufcan and Efe A. Ok (2014), "Canonical model of choice with initial endowments." *Review of Economic Studies*, 81, 851–883. [990, 991, 993, 996, 998, 1005, 1006, 1009]

Masatlioglu, Yusufcan and Neslihan Uler (2013), "Understanding the reference effect." *Games and Economic Behavior*, 82, 403–423. [990]

Mayraz, Guy (2011a), "Priors and desires." Unpublished paper, SSRN 1805564. [991, 998, 1007]

Mayraz, Guy (2011b), "Wishful thinking." Unpublished paper, SSRN 1955644. [990, 992]

Mijović-Prelec, Danica and Drazen Prelec (2010), "Self-deception as self-signaling: A model and experimental evidence." *Philosophical Transactions of the Royal Society, B*, 365, 227–240. [992]

Ortoleva, Pietro (2010), "Status quo bias, multiple priors and uncertainty aversion." *Games and Economic Behavior*, 69, 411–424. [990, 1005, 1009]

Plott, Charles R. and Kathryn Zeiler (2005), "The willingness to pay-willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations." *American Economic Review*, 95, 530–545. [1004]

Plott, Charles R. and Kathryn Zeiler (2011), "The willingness to pay—willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations: Reply." *American Economic Review*, 101, 1012–1028. [1004]

Riella, Gil and Roee Teper (2014), "Probabilistic dominance and status quo bias." *Games and Economic Behavior*, 87, 288–304. [1005, 1009]

Rubinstein, Ariel and Lin Zhou (1999), "Choice problems with a 'reference' point." *Mathematical Social Sciences*, 37, 205–209. [1009]

Saito, Kota (2013), "Social preferences under risk: Equality of opportunity versus equality of outcome." *American Economic Review*, 103, 3084–3101. [1016]

Samuelson, William and Richard Zeckhauser (1988), "Status quo bias in decision making." *Journal of Risk and Uncertainty*, 1, 7–59. [989]

Sharot, Tali (2011), "The optimism bias." *Current Biology*, 21, R941–R945. [992]

Sharot, Tali, Alison M. Riccardi, Candace Raio, and Elizabeth Phelps (2007), "Neural mechanisms mediating optimism bias." *Nature*, 450, 102–105. [992]

Spiegler, Ran (2008), "On two points of view regarding revealed preference and behavioral economics." In *The Foundations of Positive and Normative Economics: A Handbook* (Andrew Caplin and Andrew Schotter, eds.), 95–115, Oxford University Press, New York, New York. Chapter 4. [1008]

Sprenger, Charles (2015), "An endowment effect for risk: Experimental tests of stochastic reference points." *Journal of Political Economy*, 123, 1456–1499. [990, 996]

Staw, Barry M. (1976), "Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action." *Organizational Behavior and Human Performance*, 16, 27–44. [996]

Tserenjigmid, Gerelt (2019), "Choosing with the worst in mind: A reference-dependent model." *Journal of Economic Behavior and Organization*, 157, 631–652. [1009]

Weinstein, Neil D. (1980), "Unrealistic optimism about future life events." *Journal of Personality and Social Psychology*, 39, 806–820. [992]

Yariv, Leeat (2001), "Believe and let believe: Axiomatic foundations for belief dependent utility functionals." Working paper, Cowles Foundation Discussion Paper No. 1344. [1009]

Yariv, Leeat (2005), "I'll see it when I believe it: A simple model of cognitive consistency." Working paper, Cowles Foundation Discussion Paper No. 1352. [1009]

Co-editor Ran Spiegler handled this manuscript.

Manuscript received 27 December, 2016; final version accepted 26 October, 2019; available online 12 November, 2019.