Priors and Desires—A Bayesian Model of Wishful Thinking and Cognitive Dissonance*

Guy Mayraz

Department of Economics, University of Melbourne

November 14, 2018

Abstract

This paper offers a model of decision making in which beliefs depend not only on relevant evidence but also on what decision-makers want to be true. Simplifying assumptions yield an "as if" Bayesian representation, in which decision-makers observe how much each state benefits them, and take that as evidence about its likelihood. A single parameter determines both the direction and weight of this "evidence". A decision-maker's beliefs affect her subsequent decisions, which in turn determine what she wants to be true in the next period. Buying an asset causes investors to become biased about it. They will thus hold on to it despite bad news that would have prevented the purchase if the news had arrived earlier. In the asbence of bad news, investors will gradually increase their investment as they become more confident about the asset. The model applies to any domain in which beliefs are subjective. Beyond investing, I explore applications to school choice, conflict situations, redistribution, and decisions in morally ambiguous situations.

JEL classification: D01,D03,D80,D81,D83,D84.
Keywords: wishful thinking, cognitive dissonance, self-serving beliefs, optimism, pessimism

*I am grateful to Erik Eyster, Wolfgang Pesendorfer and Matthew Rabin for stimulating discussions and valuable suggestions at key points in the writing of the paper, and to seminar participants at CalTech, Collegio Carlo-Alberto, Hebrew U., LSE, Maastrict, MIT, Monash U., Princeton, Royal Holloway, Tel-Aviv U., UBC, UC Berkeley, UCL, UCSB, U. Essex, U. Melbourne, U. Sydney, UWA, and Warwick.

1 Introduction

People want many things to be true. They want their child's school to be a good one, their investments to recover, and a great many other things too. They also have beliefs about these things: *is* their child's school good? *will* their investments recover? They know, of course, that wanting something to be true does not make it so, but keeping beliefs and desires apart is not easy. Realists manage to keep their feelings walled off from their beliefs, but wishful thinkers don't. This paper offers a model of decision making that allows for wishful thinkers.

The model is outlined in Figure 1. People want certain things to be true. If they are wishful thinkers, their desires affect their beliefs, with a subsequent impact on the choices that they make. To take a simple example, the desire for good health causes wishful thinkers to underestimate risks to their health. This belief then causes them to purchase too little health insurance. Choices maximize subjective expected utility, so the departure from the standard model is limited to the impact of desires on beliefs.

The desire for good health is innate, but many desires are acquired. It is because their child has been allocated to some particular school that the parents want it to be the best school around. In the case of acquired desires, we can observe not only the resulting beliefs but also the change in beliefs brought about by the change in desires. Suppose the school allocation was originally determined by the toss of a fair coin. Before learning the outcome of the coin toss, the parents were indifferent as to which school is better, and their beliefs were not affected by wishful thinking. It is only after they learned that their child would go to school A, that they acquired the desire for A to be the better school, and consequently came to believe that it is. This change in beliefs as a result of a change in desires is an example of cognitive dissonance.

It is essential to these examples that uncertainty is subjective. The beliefs of wishful thinkers are a function of what they want to be true, but the full distribution of beliefs has the same support as that of realists (Figure 2). Individual wishful thinkers in particular situations can, therefore, always find plausible justifications for their beliefs and can continue to think of themselves as realists.

Desires are represented by a real-valued function π that maps each state to the utility of the outcome that is obtained in that state. In the school allocation example, let A denote the state in which A is the better school, let B denote the state in which B is better, and let u denote the parents' utility function, with

u(child attends the better school) $= 1$, and u(child attends the worse school) $= 0$. If the child has been allocated to school A, $\pi(A) = 1$ and $\pi(B) = 0$. If the child has been allocated to school B, $\pi(A) = 0$ and $\pi(B) = 1$. Finally, if the allocation is to be determined by the toss of a fair coin, the outcome in both states is a 50-50 lottery with an expected utility of 0.5, and so $\pi(A) = \pi(B) = 0.5$.

Subjective beliefs are represented by a probability measure p that varies with the desires π. In order to obtain a tractable model of wishful thinking, I impose several simplifying assumptions on this dependence. The most important is that p is sensitive only to how the outcome in a state compares with the outcome in other states, and not to how good it is in an absolute sense.[1] Using these assumptions, I show that there exists a probability measure p_0 and a real-valued parameter b, such that for any state s,

$$p_\pi(s) \propto p_0(s)e^{b\pi(s)}, \tag{1}$$

and for any event A,

$$p_\pi(A) \propto \int_A e^{b\pi}\,\mathrm{d}p_0. \tag{2}$$

Note first that if π is constant across all states then $p_\pi = p_0$. The probability measure p_0 represents the decision-maker's *indifference beliefs*, or the beliefs she would hold if she were indifferent between all states. More generally, p_π is distorted away from p_0, and the direction and magnitude of this distortion depend on b. If $b > 0$, p_π is distorted towards more desirable states—those in which π is relatively high. If $b < 0$, p_π is distorted in towards *less* desirable states—those in which π is relatively low. The knife edge case of $b = 0$ corresponds to *realism*: $p_\pi = p_0$ for all π. The model thus captures both wishful thinking and its pessimistic opposite and includes the standard model as a special case.

The equations of the model are formally identical to Bayes Rule, though the "evidence" is not normatively relevant. Consider beliefs conditional on evidence e and desires π:

$$\underbrace{p_\pi(s|e)}_{\text{posterior beliefs}} \propto \underbrace{p_0(s)}_{\text{prior}} \cdot \underbrace{\mathcal{L}(s|e)}_{\text{evidence}} \cdot \underbrace{e^{b\pi(s)}}_{\text{“evidence”}}\ . \tag{3}$$

[1]Formally, p is invariant to additive shifts in π. Note that this is invariance to a particular change in outcomes, and is not a trivial consequence of the flexibility in choosing the utility function.

If we identify $e^{b\pi(s)}$ with the likelihood of s given the desires π, and write $p(s)$ instead of instead of $p_0(s)$, the equation takes the exact same form as Bayes Rule:

$$p(s|e,\pi) \propto p(s) \cdot \mathcal{L}(s|e) \cdot \mathcal{L}(s|\pi). \tag{4}$$

Wishful thinkers thus behave as if the outcome that they obtain in a state is useful evidence about it. If they learn that their child has been allocated to school A, they *infer* that A is likely to be the better school. This, of course, is only an “as if” interpretation. Wishful thinkers do not actually believe that Nature chose the state of the world with their interests in mind, but they behave as if they do.

Implications Wishful thinkers overestimate the likelihood of desirable states relative to undesirable ones. The rest follows from that. Underestimating risks, they under purchase insurance, and hold on to a dangerously exposed investment portfolio. They tend to be satisfied with their own choices, and with whatever life brings them. Wishful thinkers who choose some option become biased about it. This bias then causes them to stick with their chosen option despite news that, if it had arrived before the initial decision, would have caused them to choose some other option instead.

Wishful thinkers not only stick with an option despite negative news but are also biased to increase their initial commitment if such an increase is possible. After making a risky investment, a wishful thinker will come to underestimate the risks and overestimate the benefits. Given this new assessment, she will find it optimal to increase her original investment.

The more a wishful thinker cares about something, the more biased he becomes. The more a dying wishful thinker fears death, the more likely he is to be in denial about it; the more a risk-averse an investor is, the more he underestimates the likelihood of a loss. Thus, while risk-aversion makes it less likely that an investor makes a risky investment in the first place, it also reduces the likelihood that he sells after bad news.

Wishful thinkers are biased over anything they care about. A wishful thinker who wants to be a good person will be biased to think that she is right and that the actions she performed are good and just. If a wishful thinker hit against someone in a crowded street, she is liable to think it is the other person’s fault. Tackles performed by her favorite sports team are legitimate, while those of her opponents are unfair. And of course, the more a wishful thinker values moral behavior, the more biased she becomes about the morality of her actions. A wishful thinker induced to perform a questionable act may come to believe the act is moral, and will willingly repeat it at a later opportunity.

Pessimistic wishful thinkers are biased in the opposite direction. The more they hope something is true, the less likely they are to believe it. They are dissatisfied with their choices and are predisposed to switch. A pessimistic wishful thinker who hits against someone in a crowded street will be convinced that it is her fault.

Evidence The empirical literature on wishful thinking and cognitive dissonance is vast, and cannot possibly be given justice here. To give just a couple of examples, Babcock and Loewenstein (1997) randomly assign experimental subjects to the role of either the plaintiff or the defendant in a compensation trial.[2] Subjects can agree on the compensation amount between themselves, or else the amount is determined by a judge's decision, in which case 'legal fees' are deducted from both sides. A compromise agreement is therefore efficient but is frequently thwarted by wishful thinking. Plaintiffs expect the judge to decide on a substantially higher award amount than defendants do, and this gap in expectations makes it impossible to reach a mutually acceptable compromise agreement. In an example of cognitive dissonance, Knox and Inkster (1968) elicit the beliefs of bettors who stand in line to bet in a horse race and those who already placed the bet. They find that committing to the bet increases the subjective probability that the horse would win the race.

The negative form of wishful thinking is studied primarily by psychiatrists. People suffering from depression appear to be immune to positive wishful thinking (Alloy & Abramson, 1982) and are instead biased in a negative direction (Seligman, 1998). Cognitive behavioral therapy works, in part, by challenging such unrealistically negative beliefs.

Related theory In psychology, "optimism" and "pessimism" are close cousins of wishful thinking.[3] In economics these terms refer instead to a non-linear weighting of outcomes. In rank-dependent utility models, "pessimistic" decision weights explain the Allais Paradox (Allais, 1953).[4] In models of am-

[2] Also see Loewenstein et al. (1993) and Babcock et al. (1995).

[3] Often with the added adjective "unrealistic", and typically understood as referring specifically to uncertainty about the future.

[4] Non-linear decision weights originate in Fellner (1961). Rank-dependent utility was introduced by Quiggin (1982). Wakker (1990) characterizes this notion of pessimism.

biguity aversion, a "pessimistic" attitude to ambiguity explains the Ellsberg Paradox (Ellsberg, 1961). Optimism and pessimism in these models are usually understood as either an over-confident or a cautious *attitude* to uncertainty, rather than a distortion of beliefs.[5]

An apparent exception is the Bracha and Brown (2012) model of optimism, in which distorted probabilities are interpreted as beliefs. However, as Bracha and Brown show, the resulting model is observationally indistinguishable from ambiguity seeking. Thus, a Bracha and Brown (2012) optimistic investor who considers a bet on oil prices will assign rising prices a higher probability when evaluating a bet that prices will rise than when evaluating a bet that prices will fall. This contrasts with a wishful thinking investor, who will use the same probabilities when evaluating both bets (though these probabilities will be distorted by any pre-existing stake she has in the direction that oil prices will take.)[6]

Closer to wishful thinking are models of self-deception. Decision-makers in these models have both the motive and the means to bias their own beliefs, and thus opt to deliberately deceive themselves.[7] In some models the motives for self-deception are instrumental. A decision-maker who wants others to overestimate her ability may be more convincing if she believes this herself. If she wishes her future-self to study hard for an upcoming exam, she may benefit from her future-self believing that her ability is at the point in which returns to effort are maximal (Bénabou & Tirole, 2002). In other models, decision-makers have preferences directly over their own beliefs. Workers in a hazardous industry with no safety equipment choose to believe the industry is safe—not because this belief promotes any desirable outcomes, but because they like it better than the realist alternative (Akerlof & Dickens, 1982).

Decision-makers in all these models trade off the gain in biasing their beliefs against the cost in biasing their decisions. The workers in the hazardous industry choose between complete denial of any danger and just enough fear so that they do purchase safety equipment when it later becomes available.

[5] They thus also provide an alternative interpretation of risk-seeking and risk-aversion that is compatible with a linear utility function (Dillenberger et al., 2017).

[6] Whereas a wishful thinker who gains from oil prices going up (down) behaves as if she *infers* that oil is likely to go up (down), an optimistic decision-maker behaves as if her choice of betting on oil prices going up (down) *causes* higher (lower) prices.

[7] The self-deception literature starts with Akerlof and Dickens (1982). Other particularly important contributions include Bénabou and Tirole (2002) and Brunnermeier and Parker (2005). Bénabou and Tirole (2016) provide an overview.

The more costly is denial, the more likely are decision-makers to stick with realist beliefs (or close enough to realism so that the bias makes no difference to their choices.) In some of these models, decision-makers choose beliefs freely. In others, they manipulate future beliefs through selective forgetting of negative signals. If decision-makers understand that their memory may be biased, self-deception turns into a game between multiple selves (Bénabou & Tirole, 2002).

The outcome of self-deception can resemble wishful thinking, but the details are different. Self-deception can result in extreme belief distortion, little, or none, and is unlikely to affect high-stakes decisions, in which the cost of biased beliefs exceeds any plausible enefit. Wishful thinking distorts beliefs smoothly as a predictable outcome of the decision-maker's desire for things to be true, and does so whether the cost in distorting choices is low or high. Self-deception offers a rational explanation for biased beliefs and can explain a decision to avoid information.[8] Wishful thinking offers a simpler and more general model of biased beliefs that retains the structure of instrumental rationality.

It is finally of interest to compare and contrast the present model with the Kőszegi and Rabin (2006) model of reference-dependent preferences. In both models preferences are reference-dependent, but whereas in Kőszegi and Rabin (2006) the utility function is reference-dependent, in this paper, it is beliefs that are reference-dependent. In Kőszegi and Rabin (2006) the states in which an outcome is obtained are irrelevant, but consumption dimensions are all important. In this paper, the opposite is the case. Finally, in Kőszegi and Rabin (2006) decision-makers have rational expectations over their future choices, and these expectations enter into the reference bundle. In this paper, decision-makers are naive about their future choices, and their desires are independent of the choice set. This final difference is not a coincidence—in Kőszegi and Rabin (2006) decision makers are loss-averse, and the natural modeling assumption is that they are aware of how this feature of their utility function affects their choices. In this model, decision-makers are wishful thinkers, and the natural modeling assumption is naivete.

[8]For example, in Oster et al. (2013) decision-makers avoid a test that would reveal whether they will develop Huntington Disease so that they can continue to live in denial.

2 Model

The introduction describes the overall structure of the model. This section goes into the technical details, starting with the decision-maker's subjective and objective uncertainty over the outcome.

2.1 Uncertainty

Suppose a wishful thinker stands to gain a million dollars if some business venture succeeds or if a fair coin lands on heads. Given the situation, the wishful thinker naturally wants both the business venture to succeed and the coin to land on heads. A straightforward model of wishful thinking would thus predict a bias in both cases. This, however, does not seem right: the process of belief formation over the business venture involves subjective judgment, and plausibly offers room for wishful thinking to enter; there is also likely to be a range of beliefs among similarly informed decision-makers so that the bias in a wishful thinker's beliefs will not be obvious. Neither of these is true for the coin toss.[9]

For these reasons, it would be desirable for the model to allow for wishful thinking only when uncertainty is subjective. I could, therefore, use neither the von-Neumann and Morgenstern (1944) model—in which all uncertainty is represented by lotteries with known probabilities—nor the Savage (1954) model—in which all uncertainty is represented by states. Instead, I use the Anscombe and Aumann (1963) model, which combines these two types of uncertainty in a compound lottery in which states map into lotteries with known probabilities. Let Z denote the set of outcomes that the decision-maker cares about, let S denote the set of states, and let Σ be a σ-algebra of events. An *Anscombe-Aumann mapping* is a Σ-measurable mapping $f : S \to \Delta(Z)$, associating with each state a lottery with known probabilities over the set of

[9] This is the distinction that Knight (1921) drew nearly a century ago between situations in which probabilities are assumed or calculated and those in which no such calculation is possible: "Business decisions, for example, deal with situations which are far too unique, generally speaking, for any sort of statistical tabulation to have any value for guidance. The conception of an objectively measurable probability or chance is simply inapplicable…Yet it is true, and the fact can hardly be overemphasized, that a judgment of probability is actually made in such cases."

outcomes.[10]

Example 1 (School allocation)**.** A child is to be allocated by the toss of a fair coin to one of two schools: A or B. The parents want their child to attend the better of the two schools. Let $Z = \{\text{Success}, \text{Failure}\}$ denote whether they are successful in achieving this goal, and let $S = \{A, B\}$ denote the identity of the better school. Before the allocation is decided, both states are mapped into the same 50-50 objective lottery over these outcomes: $f(A) = f(B) = (\text{Success}, 0.5; \text{Failure}, 0.5)$. Suppose now that the parents are informed that their child has been allocated to school A. This news resolves all objective uncertainty, and with some abuse of notation the mapping representing the new situation can be written as $f(A) = \text{Success}$, and $f(B) = \text{Failure}$.[11] If the child had been allocated to school B, the mapping would have been $f(A) = \text{Failure}$, and $f(B) = \text{Success}$.

2.2 Preferences, utility, and desires

Let F denote the set of all Anscombe-Aumann mappings. Elements of F can represent not only the decision-maker's uncertainty over the outcome but also objects of choice. Consider the Anscombe-Aumann mapping that in Example 1 represents the situation the parents find themselves in after their child has been allocated to school A. In a school choice setting the same mapping represents the option of sending the child to school A. Decision-makers thus have preferences over the elements of F. In order to minimize the departure from the standard model, I assume that these are subjective expected utility preferences. The only difference is that a decision-maker's beliefs can vary with what she wants to be true—her *desires*.

At any point in time, the decision-maker's uncertainty over the outcome is represented by some Anscombe-Aumann mapping f, which (among other things) determines her desires. The precise lottery in each state is not important for the desires—what matters is only how much the decision-maker values that lottery. I thus define the desires as $u \circ f$ (Definition 1). However, since the utility function is determined endogenously from preferences, Assumption 1 first asserts that the decision-maker has subjective expected utility

[10]In the Anscombe and Aumann (1963) terminology, this is a "horse lottery" with "roulette lotteries" as its prizes.

[11]Formally, the outcome in both states remains a lottery, albeit a degenerate one: $f(A) = (\text{Success}, 1; \text{Failure}, 0)$ and $f(B) = (\text{Success}, 0; \text{Failure}, 1)$.

preferences in which beliefs depend on f, and only then—once the utility function is defined—states that belief depend on f only via the associated desires $u \circ f$.

Assumption 1. Let $\succeq$ denote the decision-maker's preferences over F. There exists a utility function $u : Z \to \mathbb{R}$ and a distortion mapping $p : F \to \Delta(S)$, such that if f is the Anscombe-Aumann mapping representing the decision-maker's uncertainty over the outcome, then for any $g, g' \in F$, $g \succeq g'$ if and only if

$$\int_S u(g(s))\,\mathrm{d}p_f \geq \int_S u(g'(s))\,\mathrm{d}p_f. \tag{5}$$

Moreover, if f and f' are Anscombe-Aumann mappings such that $u \circ f = u \circ f'$ then $p_f = p_{f'}$.[12]

Definition 1 (Desires)**.** Let f be the Anscombe-Aumann mapping that represents the decision-maker's uncertainty over the outcome, and let u be her utility function. The decision-maker's *desires* is a mapping $\pi : S \to \mathbb{R}$ defined by $\pi(s) = u(f(s))$ for all $s \in S$.

Example 1 (School allocation, continued)**.** Suppose the parents' utility for the outcomes is $u(\text{Success}) = 1$ and $u(\text{Failure}) = 0$. The desires before the allocation is resolved are $\pi(A) = \pi(B) = 0.5u(\text{Success}) + 0.5u(\text{Failure}) = 0.5$. The desires after an allocation to school A are $\pi(A) = u(\text{Success}) = 1$ and $\pi(B) = u(\text{Failure}) = 0$. Before the allocation the parents are indifferent as to which school is better. After they learn that their child has been allocated to school A, they acquire the desire for A to be the better school. The same change in desires occurs if the parents choose to send their child to school A.

The utility function u can be identified independently of p by analyzing the decision-maker's preferences over Anscombe-Aumann mappings with only objective uncertainty—ones in which the objective lottery in all states is the same. This makes it possible to identify the decision-maker's desires independently of her beliefs. Let Π denote the set of all possible desires. The decision-maker's beliefs are thus given by some distortion mapping $p : \Pi \to \Delta(S)$ that associates with each desires π a Σ-measurable probability measure p over S. We now need to characterize this mapping.

[12]The two sides of Equation 5 represent the subjective expected utility of the Anscombe-Aumann mappings g and g', which are compound lotteries. Thus, $g(s)$ and $g'(s)$ are themselves lotteries (with objective probabilities), and $u(g(s))$ and $u(g'(s))$ represent the expected utility of these lotteries.

2.3 The belief distortion mapping

The distortion mapping linking beliefs to desires has to be flexible enough to model wishful thinking, but should otherwise be as simple as possible. With this goal in mind, I impose the four simplifying assumptions in Definition 2:

Definition 2 (Well-behaved distortion). $p : \Pi \to \Delta(S)$ is a *well-behaved distortion* if the following conditions are satisfied for any desires π and π', any event E, and any constant desires c:

- A1 (absolute continuity) $p_{\pi'}(E) = 0 \iff p_\pi(E) = 0$.
- A2 (prize-continuity) If $\pi_n \to \pi$ uniformly then $p_{\pi_n}(E) \to p_\pi(E)$.
- A3 (consequentialism) If $\pi = \pi'$ over a non-null[13] event E then $p_{\pi'}(A|E) = p_\pi(A|E)$ for any event $A \subseteq E$.
- A4 (shift-invariance) If $\pi' = \pi + c$ then $p_{\pi'} = p_\pi$.

The first two assumptions are technical: *Absolute Continuity* limits wishful thinking to events that the decision-maker is uncertain about, and *Prize Continuity* requires that small differences in payoffs have only a small effect on beliefs. The two substantial assumptions are *Consequentialism* and *Shift-Invariance*. According to *Consequentialism*, if two desires coincide over some event E, then the corresponding probability measures conditional on E also coincide. This assumption can be seen as the combination of Bayesian updating if E is observed with a requirement that p is affected only by desires in states that are consistent with the available evidence.[14] *Shift Invariance* requires subjective probabilities to depend only on utility differences—subjective evaluations not of the outcomes in each state as such, but of how the outcome in each state compares with the outcome in other states.

As I show in Appendix A, the combination of Absolute Continuity and Consequentialism ensures that p can be represented by a formula with the general form $p_\pi(s) \propto p_0(s)\nu(s, \pi(s))$. where p_0 is a probability measure representing beliefs under conditions of indifference (constant π), and ν is some

[13]That is, both $p_\pi(E) > 0$ and $p_{\pi'}(E) > 0$. Absolute Continuity ensures that these two requirements coincide.

[14]Since $\pi = \pi'$ on E, we must have $p_\pi(A) = p_{\pi'}(A)$ if E is observed. Assuming Bayesian updating, it follows that prior to observing E, $p_\pi(A|E) = p_{\pi'}(A|E)$.

potentially non-monotonic and discontinuous function that can also be state-dependent. The addition of Shift Invariance and Prize Continuity ensures that ν is state-independent, monotonic and continuous, and takes a particularly simple logit form: there exist some real-number b, such that $\nu(\pi(s)) = e^{b\pi(s)}$ for all s. Taken together, the four assumptions in Definition 2 are equivalent to the *Priors and Desires* formula:

Definition 3 (Priors and Desires distortion)**.** $p : \Pi \to \Delta(S)$ is a *Priors and Desires distortion* if there exists a probability measure p_0 and a real-number b, such that for any desires π and any event A,

$$p_\pi(A) \propto \int_A e^{b\pi} \, \mathrm{d}p_0. \tag{6}$$

2.4 Distorted beliefs

Equation 6 is necessary for working with probability density functions. If the state space is discrete, let $A = \{s\}$ to obtain that for any state s,

$$p_\pi(s) \propto p_0(s)e^{b\pi(s)}. \tag{7}$$

Both these equations have an implicit normalization constant, which drops out when we focus on the odds-ratio between two states:

$$\frac{p_\pi(s)}{p_\pi(s')} = \frac{p_0(s)}{p_0(s')}e^{b(\pi(s)-\pi(s'))}. \tag{8}$$

All three equations describe the wishful thinking bias relative to p_0. Beliefs coincide with p_0 if π is constant, so that the decision-maker is indifferent between all states. Even if π is not constant, beliefs coincide with p_0 if $b = 0$. The model with $b = 0$ therefore represents a standard realist decision-maker. If $b > 0$ beliefs are biased towards more desirable states, representing wishful thinking as normally understood. Finally, if $b < 0$ beliefs are biased towards *less* desirable states, providing a model of negative or pessimistic wishful thinking.

It is worth emphasizing that wishful thinkers are biased relative to what their beliefs would have been if they were disinterested observers—not relative to rational expectations. The two only coincide if wishful thinking is the *only* reason for beliefs to deviate from rational expectations.

Example 1 (School allocation, continued)**.** Suppose the parents are wishful thinkers with $b = \ln 2$, and that their indifference beliefs are $p_0(A) = p_0(B) =$

0.5. Before the allocation, the parents are indifferent between these two possibilities: $\pi(A) = \pi(B) = 0.5$, and are therefore unbiased: $p = p_0$. Since $u(\text{Success}) = 1$ and $u(\text{Failure}) = 0$, an allocation to school A changes the desires to $\pi(A) = 1$ and $\pi(B) = 0$, and so $e^{b(\pi(A)-\pi(B))} = e^{\ln 2} = 2$. Thus, according to Equation 8, wishful thinking causes the odds-ratio to double from $p_0(A)/p_0(B) = 1$ to $p_\pi(A)/p_\pi(B) = 2$. The subjective probability that A is the better school increases from $p_0(A) = 1/2$ to $p_\pi(A) = 2/3$.

2.5 Bayesian interpretation and comparative statics

The equations of the model are formally identical to Bayesian updating. If we write $p(s|\pi)$ instead of $p_\pi(s)$ and let $\mathcal{L}(s|\pi) = e^{b\pi(s)}$, Equation 7 becomes:

$$\underbrace{p(s|\pi)}_{\text{posterior beliefs}} = \underbrace{p(s)}_{\text{prior}} \cdot \underbrace{\mathcal{L}(s|\pi)}_{\text{“evidence”}} . \tag{9}$$

In this Bayesian interpretation, wishful thinkers believe that Nature is benevolent, and has chosen the state of the world with their interests in mind. If they observe that they would be better off if some state obtains, they conclude that it probably does. This, of course, is an “as if” interpretation, but it nonetheless offers some useful intuition for how the model functions. For one thing, it makes it obvious that a change in desires would cause beliefs to change. For another, it clarifies the comparative statics of the model. Since p_0 is the prior and $e^{b\pi(s)}$ is the evidence, the impact of wishful thinking on the subjective probability of a state s relative to another state s' increases with the strength of the “evidence” and with the uncertainty in the prior p_0. The more the decision-maker cares about which state is true, and the more uncertain she is, the greater the bias in her beliefs. The following examples illustrate these comparative-statics, first in the case of the odds-ratio between two states and then for a normal distribution over a continuous variable.

Example 1 (School allocation, continued)**.** We saw before that with $p_0(A) = p_0(B) = 0.5$, an allocation to school A changes beliefs to $p = 2/3$ and $p = 1/3$. Suppose now that the prior were stronger: $p_0(A) = 4/5$, so that the parents were already quite convinced to begin with that A is the better school. As before, the allocation to school A causes the odds-ratio to double (from 4 to 8), but the change in probability terms is smaller: from $p_0(A) = 4/5$ to $p_\pi(A) = 8/9$. The impact of wishful thinking also scales with the strength of the “evidence”, which in this case is $\pi(A) - \pi(B) = u(\text{Success}) - u(\text{Failure}) = 1$. If, instead, $u(\text{Success}) - u(\text{Failure}) = 2$, wishful thinking would cause the odds-ratio to quadruple, instead of merely double, leading to a distorted probability of $p_{\pi'}(A) = 16/17$.

Example 2 (Stock market)**.** A risk-neutral wishful thinker with utility function $u(x) = x$ purchases q shares for a fixed holding period. Let r denote their price at the end of the holding period, at which point the investment would be worth rq. Since $u(x) = x$, $\pi(r) = rq$. Suppose that if the investor had been indifferent, he would assign r a normal distribution with mean μ and variance σ^2. The impact of wishful thinking can be found using Equation 6:

$$
\begin{aligned}
p_\pi(r \leq r_0) &\propto \int_{r \leq r_0} e^{b\pi} \, \mathrm{d}p_0 = \frac{1}{\sqrt{2\pi}\sigma} \int_{r \leq r_0} e^{bqr} e^{-\frac{(r-\mu)^2}{2\sigma^2}} \, \mathrm{d}r \\
&\propto \int_{r \leq r_0} e^{-\frac{r^2+\mu^2-2r(\mu+bq\sigma^2)}{2\sigma^2}} \, \mathrm{d}r = \int_{r \leq r_0} e^{-\frac{(r-\mu')^2}{2\sigma^2}} \, \mathrm{d}r,
\end{aligned}
\tag{10}
$$

where $\mu' = \mu + bq\sigma^2$. Thus, while $p_0(r \leq r_0) \sim \mathcal{N}(\mu, \sigma^2)$, $p_\pi(r \leq r_0) \sim \mathcal{N}(\mu + bq\sigma^2, \sigma^2)$. The wishful thinker thus continues to believe that the asset's return is normal with variance σ^2, but instead of expecting a return of μ, he comes to expect a higher return that is shifted by $bq\sigma^2$. This bias term is increasing (i) in the level of uncertainty (as indexed by the variance σ^2), and (ii) the strength of the investor's desires, which are here determined by the number of shares that he holds.

Unlike the case with self-deception, the importance of subsequent decisions does not necessarily limit the magnitude of wishful thinking bias. Faced with a high-stakes decision, a wishful thinker would gather as much evidence as possible—just like a good Bayesian. To the extent that this information reduces uncertainty, it would also reduce the scope for wishful thinking. There are, however, many domains in which even the most informed experts are far from reaching consensus. Indifference beliefs may then remain highly uncertain, leaving plenty of scope for wishful thinking to affect beliefs and the high-stakes decisions that depend on them.

2.6 The impact of choices on subsequent beliefs

A decision made at some time t is affected by desires at t, but not vice versa (Figure 1). The decision will, however, affect desires at $t + 1$, and therefore also beliefs and decisions at $t + 1$. A decision-maker who sends the child to some school or invests in some stock acquires the desire for the school to be a good one or for the stock to do well. If she is a wishful thinker, this change in desires will affect her beliefs. Most interestingly, a decision made at t may cause beliefs at $t+1$ to change in such a way that would make some other decision optimal. For example, a risk-averse wishful thinker who invests in some asset will consequently become more optimistic about the it, and may opt to increase the original investment (Proposition 3). Some of the most interesting implications of the model follow from this multi-period dynamic.

The impact of current decisions on subsequent beliefs raises the issue of whether decision-makers take this dynamic into account in making their decisions. Consistent with the rest of the model, I assume that they do not. Decision-makers perceive themselves as Bayesian realists and make their plans accordingly. Wishful thinkers who invest in some asset assume that they will only choose to increase their investment if they obtain normatively relevant information that supports this decision. Therefore, they have no reason to forgo flexibility and commit themselves to a course of action that is optimal given their current beliefs, but which may no longer be optimal once new evidence arrives.[15]

3 Wishful thinking and cognitive dissonance

Wishful thinking and cognitive dissonance are often considered separately. In this model, however, they are the two sides of the same coin. Wishful thinking is a static relationship between beliefs and desires. Because a wishful thinker wants his health to be good, he believes his health is better than the evidence warrants. Cognitive dissonance is wishful thinking's dynamic cousin. A child is allocated to some school. Learning this causes the parents to want this particular school to be the best school one. This new desire then causes them to believe the school is better than they previously thought.

Beliefs over health offer an interesting example of the interplay between wishful thinking and cognitive dissonance. Because the desire for good health is innate and unchanging, wishful thinkers are biased about their health during their entire life. There is thus no time in which a wishful thinker was indifferent

[15]Compare models of present-biased preferences, in which sophisticated decision-makers may choose to commit themselves to some course of action despite the loss of flexibility and other costs (Laibson, 1997; O'Donoghue & Rabin, 1999).

about her health, and therefore no time in which she was unbiased. This unchanging desire can, however, attach to new objects, thereby changing beliefs about them. For example, a wishful thinker can become biased about some particular experimental drug if the drug becomes his only chance for survival. Similarly, an investor can become biased about the prospects of a business if the success of the business would make him rich, and a person looking for love can become biased about a particular individual if that individual becomes important for his love life.

3.1 *Ex-post* satisfaction with a choice

New information is one source of cognitive dissonance, but the most important source is the decision-maker's own choices. Consider a wishful thinker who starts out indifferent about the state of the world, and then either chooses some alternative A or otherwise comes to hold it. In either case, A becomes the Anscombe-Aumann mapping that represents the wishful thinker's new situation, and $\pi_A = u \circ A$ becomes her new desires. It is then natural to ask how *ex-post* satisfied is the wishful thinker with A. How does the change in desires affect the subjective expected utility of A? Does it also affect the subjective expected utility of other options? These questions are particularly interesting if the wishful thinker has an opportunity to switch from A to some other alternative. Cognitive dissonance can then affect her decision whether to take advantage of this opportunity.

Let $U_0(A)$ denote the subjective expected utility of A given the indifference beliefs p_0, and let $U_A(A)$ denote its subjective expected utility given p_π for $\pi = u \circ A$. The following proposition shows that, as long as $u \circ A$ varies between states, $U_A(A) > U_0(A)$. Wishful thinkers are thus *ex-post* satisfied with their choices. As Proposition 1 also shows, the opposite is true for negative wishful thinkers. The intuition is clear: holding A causes wishful thinkers to overestimate the likelihood of states in which A leads to desirable outcomes, and the opposite is true for negative wishful thinkers.

Proposition 1 (*Ex-post* satisfaction with choice)**.** *Let A be some Anscombe-Aumann mapping for which $u \circ A$ is not almost everywhere constant, then $U_A(A) - U_0(A)$ has the same sign as b.*

Proof. e^{bx} is increasing if $b > 0$, decreasing if $b < 0$, and constant if $b = 0$. Since $p_A(s) \propto e^{bu(A(s))}p_0(s)$ for all s, $p_A(u \circ A)$ likelihood ratio strictly dominates $p_0(u \circ A)$ if $b > 0$, is dominated by $p_0(u \circ A)$ if $b < 0$, and equals $p_0(u \circ A)$ if $b = 0$. Thus, unless A is almost everywhere constant, $U_A(A) - U_0(A)$ has the same sign of b. $\square$

Ex-post satisfaction makes wishful thinkers happy with what they have, which is great if they cannot change it. But if they *can* change it, *ex-post* satisfaction can prevent them from switching to a better option. The timing of information is thus important, even if switching is nominally costless. If a wishful thinker receives information before making a choice, she incorporates it optimally into her decision. But if she only receives the information after choosing some option, she becomes biased about it, and this bias affects her preferences. There is thus a range of weak to moderate evidence that would be enough to sway the decision if it arrives early enough, but is too weak to reverse a choice that has already been made.

Proposition 2 (Sticking with choice despite negative news)**.** *Suppose a wishful thinker who is initially indifferent between all states has to choose between two alternatives A and B, both of which are a function of states s and s', and neither of which first-order stochastically dominates the other. Then there are indifference beliefs p_0 and evidence e over s and s', such that (i) if the wishful thinker observes e before making her choice, she would strictly prefer B, and (ii) if the wishful thinker observes e only after making her choice, she would strictly prefer A both when making her choice and after observing e.*

Proof. Without loss of generality suppose $u(A(s)) > u(A(s'))$. Since neither option first-order stochastically dominates the other, $A \succeq B$ if and only if $p(s)/p(s') \geq r^*$ for some threshold r^*. Let $r_0 = p_0(s)/p_0(s')$ be the initial odds-ratio between s and s', let $r_A = e^{bu(A(s))}/e^{bu(A(s'))} > 1$ be the impact of wishful thinking if the decision-maker chooses A, and let $r_e = \mathcal{L}(s'|e)/\mathcal{L}(s|e)$ be the impact of the evidence (assumed to favor s') . Both requirements are satisfied if the following inequalities are all true:

$$r^* < r_0 < r^* r_e < r_0 r_A. \tag{11}$$

The first inequality in Equation 11 will hold whenever the indifference prior sufficiently favors A. The other two require the evidence to be strong enough that the indifference posterior favors B, but not so strong that the wishful thinking posterior also favors B. Since $r_A > 1$, it is always possible to choose r_e to satisfy both these inequalities. $\square$

One way of understanding Proposition 2 is by viewing choices as bets on events. If the decision-maker is a wishful thinker, choosing such a bet increases the subjective probability of the event in question, thereby increasing the bet's attractiveness over opposite or unrelated bets. But what about other bets on the same event? By the same reasoning, choosing one bet on some event E will cause an upwards bias in the attractiveness of other bets on E. As the following proposition shows, this process can cause a wishful thinker who starts out with a relatively small bet to increase the bet to a bigger one.

Proposition 3 (Increasing bet size)**.** *Suppose that, as in Proposition 2, a wishful thinker who is initially indifferent between all states has to choose between two alternatives A and B, both of which are a function of states s and s', and neither of which first-order stochastically dominates the other. Suppose, moreover, that both A and B are bets on s, and that B is the bigger bet:*

$$u(B(s)) > u(A(s)) > u(A(s')) > u(B(s')), \tag{12}$$

then there exist indifference beliefs p_0 and a threshold b_0 such that if $b > b_0$ the wishful thinker (i) initially prefers the smaller bet A, but (ii) after choosing A comes to prefer the bigger bet B.

Proof. It follows from Equation 12 that there is a critical threshold r^* such that the wishful thinker prefers A if and only if $p(s)/p(s') \leq r^*$. Suppose that $p_0(s)/p_0(s') = r_0 < r^*$, then the wishful thinker will initially prefer A as required. After choosing A, the odds-ratio will change to $p_A(s)/p_A(s') = e^{b(u(A(s))-u(A(s')))}r_0$. Let

$$b_0 = \frac{\ln r^* - \ln r_0}{u(A(s)) - u(A(s'))}. \tag{13}$$

For any $b > b_0$, $p_A(s)/p_A(s') = e^{b(u(A(s))-u(A(s')))}r_0 > r^*$, and so the wishful thinker will come to prefer the bigger bet B. □

4 Applications

I start this section with several applications to investing. I then consider conflicts situations and beliefs about redistribution. I end with how wishful thinkers act in situations of moral ambiguity—an important domain of decision making, albeit one rarely explored in economics papers.

4.1 Investing

Investing is a natural area for exploring the implications of wishful thinking. I start with an example of Proposition 2: an investor sticking with an investment despite bad news that would have prevented the investment if it had arrived earlier. I then show that this propensity to ignore the news varies with risk preferences. Risk-aversion increases the fear of loss and therefore leads to a greater wishful thinking bias. Thus, while risk-averse investors are less likely to make a risky investment in the first place, they can also be less likely to sell it after receiving bad news about it. Paradoxically, therefore, it can be the more risk-averse investors who end up holding a risky asset after everyone else has sold off. Investing is also a good area to illustrate the propensity for increasing bet size. In an example of Proposition 3, I show that wishful thinking can cause a risk-averse investor who prefers making only a small bet to gradually increase her bet size. I then show that this process can potentially continue until the investor is fully invested in the risky asset.

Sticking with an investment despite bad news

Proposition 2 shows that decision-makers may stick with an option despite bad news that would have caused them to make a different decision if it had arrived early enough. The following numerical example shows how this plays out in a stock market investing situation. Such a situation is particularly relevant since investors can often sell their stocks at minimal cost. Nevertheless, in what has become known as the Disposition Effect, investors appear to "ride losers too long" (Shefrin & Statman, 1985). There are, of course, many explanations for this phenomenon, but wishful thinking may also play a role.

Example 3 (Keeping a stock despite bad news)**.** An investor with indifference beliefs $p_0(A) = p_0(B) = 0.5$ decides whether to invest in company A or B. Any evidence e for which $\mathcal{L}(B|e)/\mathcal{L}(A|e) > 1$ will cause her to opt for B. But suppose that $b = \ln 2$, and that the investor only observes e after having chosen A, so that $\pi(A) = 1$ and $\pi(B) = 0$. The posterior odds-ratio after observing e is:

$$\frac{p_\pi(A)}{p_\pi(B)} = \frac{p_0(A)}{p_0(B)} \cdot \frac{e^{b\pi(A)}}{e^{b\pi(B)}} \cdot \frac{\mathcal{L}(A|e)}{\mathcal{L}(B|e)} = 1 \cdot \frac{e^{\ln 2}}{e^0} \cdot \frac{\mathcal{L}(A|e)}{\mathcal{L}(B|e)} = 2 \cdot \frac{\mathcal{L}(A|e)}{\mathcal{L}(B|e)}. \tag{14}$$

Hence, the investor will only conclude that B is better if $\mathcal{L}(B|e)/\mathcal{L}(A|e) > 2$. Weaker evidence with $1 < \mathcal{L}(B|e)/\mathcal{L}(A|e) < 2$ will no longer be sufficient.

Risk-aversion increases the required level of confidence for an investor to purchase a risky asset. But if the investor does purchase the asset, risk-aversion then increases the desire for the asset to do well, causing the investor to continue to have confidence in the asset despite bad news. As the following proposition shows, this can cause a risk-averse investor to hold on to the risky asset despite news that causes an equally informed risk-neutral investor to sell. Counter-intuitively, therefore, the people left holding to a risky asset after bad news arrive may not be those with a strong appetite for risk, but rather those who fear it.

Proposition 4. *Suppose an investor can choose at $t = 0$ whether to invest in a risky asset yielding utility 1 in state A, and utility $-\lambda$ in state B, where $\lambda \geq 1$ represents the investor's disutility for losses, and where the alternative is a safe asset yielding a utility of zero in both states. Suppose the initial evidence favors A, so that at $t = 0$, $p_0(A)/p_0(B) = \lambda_0 > 1$, but that at $t = 1$ the investor receives evidence e in favor of state B, such that a realist would come to believe B. That is, $p_0(B|e)/p_0(A|e) = \mu > 1$. Finally, suppose that after receiving the evidence the investor has an opportunity to reverse the purchase and sell the risky asset. Then there exist parameters b, λ_0 and μ, such that (i) a risk-neutral investor with $\lambda = 1$ would purchase the risky asset, but sell it after the bad news, and (ii) there exists a risk-averse investor with $\lambda > 1$, who would also purchase the risky asset, but will hold on it after receiving the bad news.*

Proof. The risky investment is optimal at $t = 0$ if $p(A) - \lambda p(B) > 0$, or

$$\frac{p(A)}{p(B)} = \lambda_0 \geq \lambda, \tag{15}$$

since $p = p_0$ at $t = 0$. If the investor makes the investment, desires change so that $\pi(A) - \pi(B) = 1 + \lambda$. The combination of the change in desires in favor of A and the evidence in favor of B results in a $t = 1$ odds-ratio of $p_\pi(A)/p_\pi(B) = e^{b(1+\lambda)}/\mu$. Given an opportunity to reverse the purchase, the investor would hold on to the risky asset if

$$\frac{p(A)}{p(B)} = \frac{e^{b(1+\lambda)}}{\mu} \geq \lambda. \tag{16}$$

Consider first the case of a risk-neutral investor, for whom $\lambda = 1$. Such an investor would buy the risky asset since $\lambda_0 > 1 = \lambda$, and would sell it after receiving the bad news if $\mu > e^{2b}$. Consider now the most risk-averse investor who purchase the asset—one for whom $\lambda = \lambda_0$. Such an investor would keep the risky asset if $e^{b(1+\lambda_0)}/\lambda_0 \geq \mu$. But as long as $b > 0$, the term $e^{b(1+\lambda_0)}/\lambda_0$ is convex in λ_0, and will therefore be true for a large enough λ_0. $\square$

Investors choosing increasingly bigger bets

Proposition 3 shows that making a small bet on some event can cause a wishful thinker to come to prefer a bigger bet. Investing is the most natural area for this to happen. The following is a numerical example:

Example 4 (Increasing investment)**.** A wishful thinking investor with log utility, $b = 1$, and wealth $w = 1$ can put some portion α of his wealth in a risky asset that doubles in state A and is worth nothing in state B. The investor thus maximizes $p(A)\ln(1+\alpha) + (1-p(A))\ln(1-\alpha)$, with the solution $\alpha = 2p(A) - 1$. If initially $p_0(A) = 2/3$, the investor chooses $\alpha = 1/3$. After making this investment, $\pi(A) - \pi(B) = \ln(4/3) - \ln(2/3) = \ln 2$. Since $b = 1$, the odds-ratio doubles, and subjective probability of the good state increases from $p_0(A) = 2/3$ to $p_\pi(A) = 4/5$. With the new beliefs, the optimal choice becomes $\alpha = 2p_\pi(A) - 1 = 3/5$. Wishful thinking thus causes the investor to increase his investment from 33.3% of his wealth to 60%.

As the following proposition shows, this process can theoretically result in the risk-averse investor putting all his money in the risky asset:

Proposition 5. *Suppose a wishful thinker with parameter $b > 0$ in the investment environment of Example 4 starts with indifference odds-ratio of $r_0 = p_0(A)/p_0(B) > 1$, and makes in each period the optimal investment decision given current beliefs. Then, (i) $p(A)$ is increasing from one period to the next, and (ii) $p(A)$ diverges ($p(A) \to 1$) if and only if $b \geq 1$.*

Proof. Let $r_k = p(A)/p(B)$ denote the odds-ratio in period k. The share invested in the risky asset in that period is

$$\alpha_k = 2p(A) - 1 = 2\frac{r_k}{r_k+1} - 1 = \frac{2r_k - (1+r_k)}{r_k+1} = \frac{r_k - 1}{r_k + 1}. \tag{17}$$

r_{k+1} can be found using Equation 8,

$$\frac{r_{k+1}}{r_0} = \frac{e^{b\pi(A)}}{e^{b\pi(B)}} = \frac{e^{b\ln(1+\alpha_k)}}{e^{b\ln(1-\alpha_k)}} = \left(\frac{1+\alpha_k}{1-\alpha_k}\right)^b = \left(\frac{2r_k}{2}\right)^b = r_k^b. \tag{18}$$

By induction, $r_k = r_0^{1+b+b^2+\cdots b^k}$, which increases from period to period and diverges if and only if $b \geq 1$. Since $p(A) = r_k/(r_k + 1)$, $p(A)$ is also increasing from period to period, and $p(A) \to 1$ if and only if $b \geq 1$. $\square$

It is essential to Proposition 5 that the investor can believe that the risky asset is completely safe. More realistically, the investor will consider several sources of risk—not all of which are subjective. In this more realistic setting, even the most wishful thinking investor will not believe that the risky asset is completely safe, and will therefore always keep some of his money in a safer place.

4.2 Conflict

There are many situations in which people find themselves on one side or another of a zero-sum competition or conflict. War is the most dramatic example, but there are many others: political competition, competitive sports, litigation, and many other situations. In all of these, people naturally want their side to win. Moreover, since people also care about justice, they want their side to *deserve* to win. To the extent that people are wishful thinkers, the prediction is straightforward: both sides will typically expect to win, and will also believe that justice is on their side. The implication for choices is also straightforward: an inefficient continuation of the conflict. Given their respective beliefs, both sides expect victory. Not only that, but a continuation of the fight is perceived to be the morally right thing to do.

Babcock and Loewenstein (1997) study just such a conflict situation in an experiment on civil litigation. They indeed find that both sides are biased both about their chances of winning the conflict and about the fairness of their side's argument. They are, moreover, able to show that this bias in beliefs has costly consequences. Because of the wishful thinking bias, there are many pairs of subjects that are far apart in their beliefs. Such pairs of subjects are much less likely to reach an efficient pre-trial agreement that avoids expensive legal costs.

4.3 Redistribution

As a rule, poor people support redistribution and other policies that favor the poor, while the rich oppose these policies. These policy preferences can, of course, be the result of naked self-interest, but it is a straightforward prediction of the model that they would be supported by associated beliefs that the policies are right for society as a whole. This application focuses on redistribution, but the same logic extends to many other political conflicts.

Consider two wishful thinkers: Alice and Bob. Alice is poor and stands to gain from redistribution, while Bob is wealthy, and would lose out from redistribution. Let A be the state in which a particular redistribution policy would increase economic efficiency (perhaps by giving poor people the resources to find a better job, or open a business), and let B be the state in which it would make things worse (perhaps by reducing incentives to work). Alice's desires are higher in A: $\pi_A(A) > \pi_A(B)$, whereas Bob's desires are higher in B: $\pi_B(B) > \pi_B(A)$. Hence, by Equation 8, Alice is biased to believe that redistribution would enhance economic efficiency, whereas Bob is biased to believe that it would make things worse. A similar result can, of course, be obtained about the policy's fairness, or the relative importance of luck in effort in determining economic outcomes.

4.4 Moral ambiguity

People often want to do the morally right thing, but many situations are morally ambiguous. When a decision-maker who cares about morality performs some action, she acquires the desire for it to be the morally right thing to do. If she is a wishful thinker, she would be biased to believe that it is.

Consider wishful thinkers who are given some material inducement to perform a morally questionable action. If the inducement is sufficient, they may perform the action despite its likely immorality. Having performed the action, they will become biased about it and may end up believing that the action is morally right. As the following example shows, this requires them to care just enough about morality: too much and the inducement will not be enough to get them to perform the action; too little and they will perform the action readily, but remain unbiased; just enough and they will reluctantly agree to perform the action, and then resolve their inner conflict by coming to believe that they did nothing wrong.

Example 5 (Morally ambiguous action)**.** A wishful thinker with $b = \ln 2$ has utility $y + \alpha m$ over material rewards y and the morality m of an action, with $\alpha \geq 0$ being a parameter that measures the importance of morality relative to money. In state A the action is moral, with $m = 1$, and in state B it is immoral, with $m = -1$. If the decision-maker does choose to perform the action, he gets a utility of $\pi(A) = 1 + \alpha$ in state A, and $\pi(B) = 1 - \alpha$ in state B. Thus, $\pi(A) - \pi(B) = 2\alpha$. Since $b = \ln 2$, the distortion factor in Equation 8 is 4^{α}. Conditional on performing the action, therefore, the more the wishful thinker cares about morality, the greater the impact of wishful thinking on her beliefs.

Suppose that decision-makers are offered an inducement $y = 1$ to perform the action, and that $p(A) = 1/9$. Utility is $1 + \alpha/9 - (8/9)\alpha = 1 - 7\alpha/9$. The decision-maker will thus perform the action if $\alpha < 9/7$. But since $p(A)/p(B) = 1/8$, decision-makers with $\alpha \in (1, 9/7)$ will have posterior beliefs of $p_{\pi}(A) > 1/2$ after performing the action. They will thus be willing to do the action a second time even without any material inducement.

5 Negative wishful thinking

Negative or pessimistic wishful thinking is the mirror image of positive wishful thinking. Whereas positive wishful thinkers believe what they want to be true, negative wishful thinkers believe the opposite. In the school allocation example, negative wishful thinkers would come to believe that their child has been allocated to the worse school. In the investing examples, they would come to believe that they invested in a bad stock. Thus, whereas positive wishful thinkers are biased to stick with what they have, negative wishful thinkers are biased to switch. As the following example shows, they may even end up in a cycle of indecision:

Example 6 (Indecision)**.** A negative wishful thinker with $b = -\ln 2$ has to decide between betting on oil prices moving up on down. After making the initial choice, the wishful thinker has unlimited opportunities to costlessly change his mind. Prices go up in state U and down in D. The payoff in utility terms is 1 if betting correctly, and 0 if betting incorrectly. Suppose that $p_0(U)/p_0(D) = r_0 \geq 1$, so the initial bet is on rising prices. According to Equation 8, the odds-ratio then changes to $r_U = e^{-\ln 2} r_0 = r_0/2$. Hence, if $r_0 \geq 2$, the negative wishful thinker sticks with the bet on rising prices, but if $1 \leq r_0 < 2$, the negative wishful thinker will switch to a bet on falling prices. In that case, the odds-ratio will change again to $r_D = 2r_0 > 1$. Hence the negative wishful thinker will switch yet again to betting on rising prices and will continue to switch back and forth until evidence arrives that puts the indifference odds-ratio either above 2 or below 1/2.

6 Conclusion

Wishful thinking can affect beliefs whenever subjective judgment is involved—a vast category that includes most of the situations economists are typically interested in, and many others besides. Of course, it remains an open question how important this impact really is. It is a key implication of the model that wishful thinking affects high stakes decisions no less than small stakes ones. Whether this is indeed the case is an important question for empirical research.

The model developed in this paper can be readily used to study the impact of wishful thinking in any existing application in which decision-makers maximize subjective expected utility. In particular, a rational expectations assumption can be replaced with the assumption that p_0 corresponds with rational expectations. Wishful thinking would then be the only reason for beliefs to deviate from rational expectations.

In this paper, I focused on reduced-form models, mostly with only two states: a good state and a bad one. There are plenty of applications for which such models are sufficient, but there are cases in which more elaborate models can provide further insights. When a number of variables enter the utility function as substitutes or complements, news about one can change the desire for another. Modeling such effects requires a more complicated state-space. A very different reason is to study the impact of wishful thinking on memory and the perceived reliability of evidence.[16]

The *Priors and Desires* formula is highly restrictive, having just one degree of freedom. A natural generalization would be to replace $e^{b\pi(s)}$ with an arbitrary strictly monotonic function—increasing for wishful thinkers and decreasing for pessimistic wishful thinkers. I have chosen not to do so here for two reasons. As I argue in Section 2.3, allowing for more degrees of freedom would reduce the predictive power of the model. But in addition, the utility function already maps outcomes into utils, and it is not clear we want to introduce a second arbitrary mapping. Consider money outcomes. A risk-averse investor has a concave utility function over money, which in the *Priors and Desires* formula ensures that the possibility of a large loss has more of an impact on beliefs than an equal sized gain. If $e^{b\pi(s)}$ is replaced with an arbitrary increasing function, a given monetary gain could potentially have more of an

[16]Whereas in some models of self-deception biased memory causes the bias in beliefs, a wishful thinker's memory is biased because of the underlying bias in beliefs that are related to this memory. The school to which a child has been allocated to will determine their beliefs about its quality, which will in turn affect how they remember their visit to the two schools.

impact on beliefs than an equal sized loss. It is not at all clear that this additional flexibility would improve the model.

One weakness of the model relative to models of self-deception is that it offers no rationale for why decision-makers may be biased—whether in a positive direction or a negative one. The cost of wishful thinking is clear enough, but what—if anything—is the benefit? One option is to dodge this question entirely, and argue that one's role as a modeler is limited to offering a mathematical description of behavior. But I also think that while wishful thinking is often harmful on a case-by-case basis, it may nonetheless have an evolutionary rationale. Recall that both positive and negative self-deception can, on occasion, be optimal on instrumental grounds. Suppose that people cannot choose to bias their beliefs on a case-by-case basis, but that evolution can nevertheless choose a person's overall level of wishful thinking (which in this model amounts to their wishful thinking parameter b). There can then be an evolutionary equilibrium consisting of some mix of positive and negative wishful thinkers.

References

Akerlof, G., & Dickens, W. (1982). The Economic Consequences of Cognitive Dissonance. *American Economic Review*, *72*(3), 307–319.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica*, *62*, 503–546.

Alloy, L. B., & Abramson, L. Y. (1982). Learned helplessness, depression, and the illusion of control. *Journal of Personality and Social Psychology*, *42*(6), 1114.

Anscombe, F., & Aumann, R. (1963). A Definition of Subjective Probability. *Annals of Mathematical Statistics*, *34*(1), 199–205.

Babcock, L., & Loewenstein, G. (1997). Explaining Bargaining Impasse: The Role of Self-Serving Biases. *Journal of Economic Perspectives*, *11*(1), 109–126.

Babcock, L., Loewenstein, G., Issacharoff, S., & Camerer, C. (1995). Biased Judgments of Fairness in Bargaining. *American Economic Review*, *85*, 1337–1343.

Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics*, *117*(3), 871–915.

Bénabou, R., & Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, *30*(3), 141–164.

Bracha, A., & Brown, D. (2012). Affective decision making: A theory of optimism bias. *Games and Economic Behavior*.

Brunnermeier, M., & Parker, J. (2005). Optimal Expectations. *American Economic Review*, *95*(4), 1092–1118.

Dillenberger, D., Postlewaite, A., & Rozen, K. (2017). Optimism and pessimism with expected utility. *Journal of the European Economic Association*, *15*(5), 1158–1175.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, 643–669.

Fellner, W. (1961). Distortion of subjective probabilities as a reaction to uncertainty. *Quarterly Journal of Economics*, 670–689.

Knight, F. (1921). *Risk, Uncertainty and Profit*. Houghton Mifflin.

Knox, R., & Inkster, J. (1968). Postdecision dissonance at post time. *Journal of Personality and Social Psychology*, *8*(4), 319–23.

Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, *121*(4), 1133–1165.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, *112*(2), 443–447.

Loewenstein, G., Issacharoff, S., Camerer, C., & Babcock, L. (1993). Self-serving assessments of fairness and pretrial bargaining. *Journal of Legal Studies*, *22*(1), 135–159.

O'Donoghue, T., & Rabin, M. (1999). Doing It Now or Later. *American Economic Review*, *89*(1), 103–124.

Oster, E., Shoulson, I., Dorsey, E. R., et al. (2013). Optimal expectations and limited medical testing: Evidence from huntington disease. *American Economic Review*, *103*(2), 804–30.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, *3*(4), 323–343.

Savage, L. J. (1954). *The Foundations of Statistics*. New-York: Wiley.

Seligman, M. (1998). *Learned Optimism*. Free Press.

Shefrin, H., & Statman, M. (1985). The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence. *Journal of Finance*, *40*(3), 777–790.

von-Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Wakker, P. (1990). Characterizing optimism and pessimism directly through comonotonicity. *Journal of Economic Theory*, *52*(2), 453–463.

A Representation theorem

Before stating the representation theorem formally, it is necessary to introduce the following technical definition:

Definition 4 (Minimally complex distortion)**.** $p : \Pi \to \Delta(S)$ is *minimally complex* if there exists desires π and three disjoint events A, B, and C, such that $p_\pi(A)$, $p_\pi(B)$, and $p_\pi(C)$ are all positive.

This requirement for there to be at least three disjoint non-null events is necessary for the Consequentialism assumption to have bite. It is not a significant restriction, since it is always possible to make a state-space more complicated by splitting states along some additional dimension.[17]

Theorem 1 (Representation theorem)**.** *A minimally complex distortion is a Priors and Desires distortion if and only if it is well-behaved.*

A.1 Finitely many events

I first prove Theorem 1 for the special case of an algebra containing finitely many events. That is, I assume that there exists a finite partition $\mathcal{S}$ of the state-space, such that Σ is the algebra generated by $\mathcal{S}$ (this includes—but is not limited to—the case where the state-space is itself finite). In addition, I prove a sequence of partial representation results requiring fewer than the four assumptions in the definition of a well-behaved distortion. In order to state the necessary and sufficient conditions for these representations I define a new

[17]The following counter-example demonstrates that the minimal complexity assumption is necessary for Theorem 1. Let $\mathcal{S} = \{A, B\}$, and define a distortion π by

$$p_\pi(A) = \frac{(\pi(A) - \pi(B))^2 + 1}{(\pi(A) - \pi(B))^2 + 2} \quad \text{and} \quad p_\pi(B) = \frac{1}{(\pi(A) - \pi(B))^2 + 2} \tag{19}$$

This distortion satisfies Absolute Continuity and Consequentialism. Nevertheless, the odds ratio between A and B cannot be expressed as the ratio of two expressions, such that the first depends only on $\pi(A)$ and the second only on $\pi(B)$, as would be required by the representation in Equation 21, let alone Equation 23.

property, *Indifference*, which is related to shift-invariance, but is considerably weaker:

A3' (Indifference). $p_\pi = p_{\pi'}$ if both π and π' are constant desires.

Note that Indifference does not require the set of utilities to have cardinal (or even ordinal) meaning. With Indifference defined, the claim consisting of the partial representation results can be stated. In all the four parts of Lemma 1 the proof that the requirements are necessary is trivial. I thus prove only that the requirements are sufficient.

Lemma 1. *Suppose that there exists a finite partition $\mathcal{S}$ of the state-space, such that Σ is the algebra generated by $\mathcal{S}$, and that π is minimally complex, then:*

1. *Absolute Continuity is a necessary and sufficient condition for there to exist a probability distribution $p_0 \in \Delta$ and a function $h : \Pi \times \mathcal{S} \to \mathbb{R}_+$, such that for any desires π and any event $A \in \mathcal{S}$,*[18]

$$p_\pi(A) \propto p_0(A) h_\pi(A). \tag{20}$$

2. *Consequentialism is a necessary and sufficient additional condition for there to exist a probability distribution $p_0 \in \Delta$, and a mapping $\mu : \mathcal{S} \times X \to \mathbb{R}_+$, such that for any desires π and any event $A \in \mathcal{S}$,*

$$p_\pi(A) \propto p_0(A) \mu_A(\pi(A)). \tag{21}$$

3. *Indifference is a necessary and sufficient additional condition for there to exist a probability distribution $p_0 \in \Delta$, and a mapping $\nu : X \to \mathbb{R}_+$, such that for any desires π and any event $A \in \mathcal{S}$,*

$$p_\pi(A) \propto p_0(A) \nu(\pi(A)). \tag{22}$$

4. *Shift-Invariance and Prize-Continuity are necessary and sufficient additional conditions for there to exists a probability distribution $p_0 \in \Delta(S)$, and a parameter $b \in \mathbb{R}$, such that for any desires π and any event $A \in \mathcal{S}$,*

$$p_\pi(A) \propto p_0(A) e^{b\pi(A)}. \tag{23}$$

[18] This is a special case of the Radon-Nikodym theorem.

Proof of Part 1. Let a denote some arbitrary constant desires. Define $p_0 = p_a$, and let $\mathcal{S}^* = \{A \in S : p_0(A) > 0\}$. Define $h_\pi(A) = p_\pi(A)/p_0(A)$ for $A \in \mathcal{S}^*$ and $h_\pi(A) = 0$ for $A \notin \mathcal{S}^*$. For $A \in \mathcal{S}^*$ the claim follows from the definition of h_π. By Absolute Continuity $p_0(A) = 0 \Rightarrow p_\pi(A) = 0$, and hence the claim holds also for $A \notin \mathcal{S}^*$. □

Proof of Part 2. Let $A \in \mathcal{S}^*$ and $x \in X$, let $\pi(A, x)$ be the desires mapping A to x and all states outside A to a. Let $E_1, \ldots, E_n$ denote the other events in $\mathcal{S}^*$. By Minimal Complexity and Absolute Continuity $\mathcal{S}^*$ includes at least two events other than A. $\pi(A, x)$ and the constant desires a agree on E_i and E_j for all i and j. Hence, by Consequentialism with $E = E_i \cup E_j$, $p_{\pi(A,x)}(E_i)/p_{\pi(A,x)}(E_j) = p_0(E_i)/p_0(E_j)$. Thus,

$$1 - p_{\pi(A,x)}(A) = \sum_i p_{\pi(A,x)}(E_i) = \sum_i \frac{p_{\pi(A,x)}(E_j)}{p_0(E_j)} p_0(E_i) = \frac{p_{\pi(A,x)}(E_j)}{p_0(E_j)} (1 - p_0(A)) \tag{24}$$

Define $\mu_A(x) = \left(\frac{1-p_0(A)}{p_0(A)}\right)\left(\frac{p_{\pi(A,x)}(A)}{1-p_{\pi(A,x)}(A)}\right)$. By Equation 24,

$$p_0(A)\mu_A(\pi(A)) = (1-p_0(A))\frac{p_{\pi(A,\pi(A))}(A)}{1 - p_{\pi(A,\pi(A))}(A)} = p_0(E_j) \cdot \frac{p_{\pi(A,\pi(A))}(A)}{p_{\pi(A,\pi(A))}(E_j)} \tag{25}$$

Let π be any desires, and let A and B be any two events in $\mathcal{S}^*$. Let π' be a desires that coincides with π on A and B, and with a elsewhere, and let C be any third event in $\mathcal{S}^*$.[19] Inserting $E_j = C$ in Equation 25 we obtain that

$$\begin{aligned} \frac{p_\pi(A)}{p_\pi(B)} &= \frac{p_{\pi'}(A)}{p_{\pi'}(B)} = \frac{p_{\pi'}(A)/p_{\pi'}(C)}{p_{\pi'}(B)/p_{\pi'}(C)} \\ &= \frac{p_0(C)}{p_0(C)} \cdot \frac{p_{\pi(A,\pi(A))}(A)/p_{\pi(A,\pi(A))}(C)}{p_{\pi(B,\pi(B))}(B)/p_{\pi(B,\pi(B))}(C)} = \frac{p_0(A)\mu_A(\pi(A))}{p_0(B)\mu_B(\pi(B))} \end{aligned} \tag{26}$$

where the first and third steps follows from Consequentialism, and the final step from Equation 25. Since Equation 26 holds for all $A, B \in \mathcal{S}^*$ it follows that Equation 21 holds for any event $A \in \mathcal{S}^*$. For an event $A \notin \mathcal{S}^*$, define $\mu_A(x) = 1$. Since $p_\pi(A) = p_0(A) = 0$ for $A \notin \mathcal{S}^*$ Equation 21 holds however μ_A is defined. Combining these results Equation 21 holds for any desires π and any event $A \in \mathcal{S}$. □

[19] This is where Minimal Complexity is used. It is essential that a third event C can be chosen.

Proof of Part 3. Let $A^* \in \mathcal{S}^*$ be some event. Define the mapping $\nu : X \to \mathbb{R}_+$ by $\nu(x) = \mu_{A^*}(x)$. For $x \in X$ let x denote also the constant desires yielding the utility x in all states. Inserting $\pi = x$ and $B = A^*$ in Equation 26 we obtain that for all $A \in \mathcal{S}^*$ and $x \in X$,

$$\frac{p_x(A)}{p_x(A^*)} = \frac{p_0(A)}{p_0(A^*)} \cdot \frac{\mu_A(x)}{\nu(x)} \tag{27}$$

Since x is a constant desires it follows from Indifference that $p_x = p_a = p$. Hence, $\mu_A(x) = \nu(x)$. Thus, $p_\pi(A) \propto p_0(A) \cdot \nu(\pi(A))$ for all $A \in \mathcal{S}^*$. Finally, this is also trivially true for $A \notin \mathcal{S}^*$, since $p_\pi(A) = p_0(A) = 0$ for $A \notin \mathcal{S}^*$. □

Proof of Part 4. Let $A, B \in \mathcal{S}^*$ be two events, and let x and y be real-numbers such that x, y, and $x + y$ are in X. Define the desires p_x and $\pi'_{x,y}$ as follows: $p_x(s) = x$ for $s \in A$ and $p_x(s) = 0$ for $s \notin A$, and $\pi'_{x,y} = p_x + y$. By Shift-Invariance, $p_{\pi'_{x,y}} = p_{p_x}$, and in particular $p_{\pi'_{x,y}}(A)/p_{\pi'_{x,y}}(B) = p_{p_x}(A)/p_{p_x}(B)$. By Equation 22 it follows that $\nu(x + y)/\nu(y) = \nu(x)/\nu(0)$. Hence, defining $\sigma(x) = \log(\nu(x)/\nu(0))$ we obtain that σ is linear, i.e. for all x and y,

$$\sigma(x + y) = \sigma(x) + \sigma(y) \tag{28}$$

For $m \in \mathbb{N}$ let $y = mx$ in Equation 28. By induction we obtain that $\sigma(mx) = m\sigma(x)$. Similarly, for $n \in \mathbb{N}$ let $y = x/n$ to obtain that $\sigma(x) = \sigma(ny) = n\sigma(y)$, and hence $\sigma(x/n) = \sigma(x)/n$. Let $y = -x$ to obtain that $\sigma(-x) = -\sigma(x)$. Combining these results, and defining $b = \sigma(1)$, we obtain that for any rational number $q \in X$, $\sigma(q) = bq$, and so $\nu(q) = \nu(0)e^{bq}$. Let now $x \in X$ be any feasible utility-value, and let $\{q_n\}_{n\in\mathbb{N}}$ be a sequence of rational feasible utility-values converging to x. By prize-continuity $p_{\pi_{q_n}} \to p_{\pi_x}$, which given Equation 22 implies that $\nu(q_n) \to \nu(x)$. By the result for rational numbers, $\nu(q_n) = \nu(0)e^{bq_n}$, and hence $\nu(q_n) \to \nu(0)e^{bx}$. Thus, $\nu(x)$ and $\nu(0)e^{bx}$ are both the limit of the same sequence of real-numbers, and so $\nu(x) = \nu(0)e^{bx}$. Finally, since Equation 22 is invariant to multiplying ν by a positive number, we obtain that Equation 23 holds for all $x \in \mathbb{R}$. □

A.2 The general case

The first step is to generalize Equation 23 to any desires and any constant utility events:

Lemma 2. *Suppose $p : \Pi \to \Delta(S)$ is a minimally complex well-behaved distortion then there exists a probability measure p_0, and a parameter $b \in \mathbb{R}$, such that for any desires π and any events A and B for which π is constant on A and B and $p_0(B) > 0$,*

$$\frac{p_\pi(A)}{p_\pi(B)} = \frac{p_0(A)}{p_0(B)} \cdot \frac{e^{b\pi(A)}}{e^{b\pi(B)}} \tag{29}$$

Proof. Let $a \in \Pi$ denote some constant desires, and define $p = p_a$. By Minimal Complexity and Absolute Continuity there exists a finite partition $\mathcal{S}$ of the state-space consisting of at least three events, such that $p_\pi(A) > 0$ for any $\pi \in \Pi$ and $A \in \mathcal{S}$. Let $\Sigma(\mathcal{S}) \subseteq \Sigma$ denote the algebra generated by $\mathcal{S}$, and let $\Pi(\mathcal{S}) \subseteq \Pi$ denote the set of $\Sigma(\mathcal{S})$-measurable desires. By Lemma 1 there exists a probability measure $p_{\mathcal{S}}$ over $(S, \Sigma(\mathcal{S}))$ and a parameter $b_{\mathcal{S}} \in \mathbb{R}$ such that Equation 23 holds any probability measure $\pi \in \Pi(\mathcal{S})$ and any event $A \in \mathcal{S}$. In particular $a \in \Pi(\mathcal{S})$ (any constant desires is), and hence for any $A \in \mathcal{S}$, $p_0(A) = p_a(A) \propto p_{\mathcal{S}}(A)e^{ba}$. Thus, $p_0(A) = p_{\mathcal{S}}$ for any event $A \in \mathcal{S}$, and hence also for any event $A \in \Sigma(\mathcal{S})$. Define $b = b_{\mathcal{S}}$. It follows that for any desires $\pi \in \Sigma(\mathcal{S})$ and any event $A \in \mathcal{S}$, $p_\pi(A) \propto p_0(A)e^{b\pi(A)}$.[20]

Let now A and B denote any events that $p_0(B) > 0$, and let π be any desires. I need to show that Equation 29 holds. To simplify notation let $\delta_\pi(A, B) = \log p_\pi(A)/p_\pi(B) - \log p_0(A)/p_0(B)$. With this notation I need to prove that $\delta_\pi(A, B) = b(\pi(A) - \pi(B))$. Let $E_1, E_2, \ldots E_n$ denote the events in $\mathcal{S}$. Without limiting generality suppose $A \cap E_1$ is not-null. Define a desires $\pi' \in \Pi$ by $\pi' = \pi(A)$ on $A \cap E_1$ and $\pi' = \pi(B)$ elsewhere, and a desires $\pi'' \in \Pi(\mathcal{S})$ by $h = \pi(A)$ on E_1 and $h = \pi(B)$ elsewhere. With these definitions,

$$\begin{aligned}\delta_\pi(A, B) &= \delta_\pi(A \cap E_1, B) = \delta_{\pi'}(A \cap E_1, B) = \delta_{\pi'}(A \cap E_1, B \cup E_2) \\ &= \delta_{\pi'}(A \cap E_1, E_2) = \delta_{\pi''}(A \cap E_1, E_2) = \delta_{\pi''}(E_1, E_2) = b(\pi(A) - \pi(B))\end{aligned} \tag{30}$$

where the last step uses the fact that π'' is in $\Pi(\mathcal{S})$, and the other steps use Consequentialism and the fact that by Shift-Invariance $p_{\pi(A)} = p_{\pi(B)} = p_0$. $\square$

Theorem 1 for the subset of simple desires is an immediate corollary.[21] The following claim is a little more general, allowing for functions that are almost everywhere simple:

[20] Note that $p = p_a$ is a probability measure over *all* the events in Σ—not just the events in $\Sigma(\mathcal{S})$.

[21] Desires π are simple if $\pi(S)$ is finite.

Definition 5. Desires $\pi \in \Pi$ are *almost everywhere simple* if there exists desires $\pi' \in \Pi$ and an event E such that π obtains only finitely many values on E and $p_{\pi'}(E) = 1$.

Corollary 1. *Theorem 1 holds when restricted to desires that are almost everywhere simple.*

Proof. The proof that a *Priors and Desires* distortion is well-behaved is trivial. I thus prove only that if π is well-behaved then it is a *Priors and Desires* distortion. Suppose π is a.e. simple then there exist a finite set of disjoint events $\{E_1, \ldots, E_n\}$ such that π is constant on any of these events, and for some desires g, $p_{\pi'}(\cup_i E_i) = 1$. By Absolute Continuity also $p_\pi(\cup_i E_i) = 1$, and so $p_\pi(A \cap \cup_i E_i) = p_\pi(A)$. Given that the events are disjoint it follows that $p_\pi(A) = \sum_i p_\pi(A \cap E_i)$. Using Lemma 2 we obtain that $p_\pi(A) \propto \sum_i p_0(A \cap E_i) e^{b\pi(A \cap E_i)}$. By Absolute Continuity $p_0(S \setminus \cup_i E_i) = 0$, and hence $\int_A e^{bf} \, \mathrm{d}p = \sum_i p_0(A \cap E_i) e^{b\pi(A \cap E_i)}$. Combining these observations we obtain that $p_\pi(A) \propto \int_A e^{b\pi} \, \mathrm{d}p_0$. $\square$

The remaining case involves functions which are *not* almost everywhere simple. If such desires exist there must exist an infinite sequence of non-null events $\{A_n\}_{n \in \mathbb{N}}$. But then, as long as $b \neq 0$ and the set of feasible utility values is unbounded, it is possible to construct desires π such $\lim_{n \to \infty} p_\pi(A_n)/p_\pi(A_1) = \infty$. But this implies that $p_\pi(A_1) = 0$, contradicting Absolute Continuity. Hence, if $b \neq 0$ the set of feasible utility values must be bounded:

Lemma 3. *Suppose $p : \Pi \to \Delta$ is a minimally complex well-behaved distortion, and that there exists desires π that is not everywhere simple, then there exist an upper bound $M \in \mathbb{R}$, such that for any feasible utility-value x, $e^{bx} \leq M$.*

Proof. The case of $b = 0$ is trivial. Henceforth I assume $b \neq 0$. By Corollary 1 there exist a probability measure p_0 and a parameter b such that Equation 32 holds for any desires π that is almost everywhere simple. If there exists a desires π that is not almost everywhere simple then there exists an infinite sequence $\{A_n\}_{n \in \mathbb{N}}$ of disjoint non-null events.[22] I need to prove that in this

[22] If π has infinitely many atoms these atoms can form the sequence. Otherwise, let E denote the event outside the set of atoms (if any). E cannot be null, or else π is almost always simple. Since π has no atoms on E it follows that there exists a value y (the median of π on E) such that $p_0(s \in E : \pi(s) \leq y) = p_0(E)/2$. Thus E includes two non-null events on which π has no atoms: $E(y)$ and $E \setminus E(y)$. This process can be repeated recursively, where in the n'th stage E is split into 2^n disjoint non-null events. An infinite sequence of disjoint non-null events can therefore be formed.

case there exists a number $M \in \mathbb{R}$ such that $e^{bx} \leq M$ for all $x \in X$. Suppose otherwise, then it is possible to choose from X a sequence $\{x_n\}_{n\in\mathbb{N}}$, s.t. for all n, $p_0(A_n)e^{bx_n} \geq p_0(A_1)e^{bx_1}$. Define a desires π by $\pi(A_n) = x_n$ for $n \in \mathbb{N}$, and $\pi(s) = x_1$ outside $\cup_n A_n$. For $n \in \mathbb{N}$ define also a simple desires π_n by $\pi_n(A_n) = x_n$ and $\pi_n(s) = x_1$ for $s \notin A_n$. By construction π and π_n agree on A_1 and A_n. The following inequality therefore holds for all $N \in \mathbb{N}$,

$$\begin{aligned}1 &\geq \sum_{n\leq N} p_\pi(A_n) = p_\pi(A_1)\sum_{n\leq N}\frac{p_\pi(A_n)}{p_\pi(A_1)} = p_\pi(A_1)\sum_{n\leq N}\frac{p_{\pi_n}(A_n)}{p_{\pi_n}(A_1)} \\ &= p_\pi(A_1)\sum_{n\leq N}\frac{p_0(A_n)e^{bx_n}}{p_0(A_1)e^{bx_1}} \geq p_\pi(A_1)\sum_{n\leq N} 1 = Np_\pi(A_1)\end{aligned} \tag{31}$$

where the second equality follows from Consequentialism, and the third from Corollary 1. Letting $N \to \infty$ we obtain that $p_\pi(A_1) = 0$, which is a contradiction to the assumption that A_1 is not null. □

Lemma 3 ensures that e^{bX} is bounded from above. If it is also bounded from below then a limit argument using simple desires can be used to extend the claim further:

Lemma 4. *Suppose $p : \Pi \to \Delta(S)$ is a minimally complex well-behaved distortion then there exists a probability measure p_0, and a parameter $b \in \mathbb{R}$, such that for any events A and B for which $p_0(B) > 0$, and any desires π for which there exist a number $m > 0$ such that $\pi(s) \geq m$ for all $s \in A \cup B$,*

$$\frac{p_\pi(A)}{p_\pi(B)} = \frac{\int_A e^{b\pi}\, dp}{\int_B e^{b\pi}\, dp} \tag{32}$$

Proof. By Corollary 1 there exist a probability measure p_0 and a parameter b such that Equation 32 holds for any desires π that is almost everywhere simple. I need to show that the claim holds also for π that is not everywhere simple. If such a desires exists then by Lemma 3 there exists a number M such that $e^{bx} \leq M$ for all $x \in X$. I assume that $b \neq 0$, since the claim is trivially true if $b = 0$. For any $n \in N$ divide the interval $[m, M]$ into 2^n non-overlapping intervals of length $(M - m)/2^n$. For any state s let $I_n(s)$ denote the interval to which e^{bf} belongs, and let $I_n^{min}(s)$ denote its lower endpoint. Define a simple desires π_n by $\pi_n(s) = \log I_n^{min}(s)/b$. With this definition $\pi_n \to f$ uniformly. In addition, $e^{b\pi(s)} - (M-m)/2^n \leq e^{b\pi_n(s)} \leq e^{b\pi(s)}$ for all s, and so $e^{b\pi_n(s)} \nearrow e^{b\pi(s)}$ for all s. Thus,

$$\begin{aligned}\frac{p_\pi(A)}{p_\pi(B)} &= \frac{\lim_{n\to\infty} p_{\pi_n}(A)}{\lim_{n\to\infty} p_{\pi_n}(B)} = \lim_{n\to\infty} \frac{p_{\pi_n}(A)}{p_{\pi_n}(B)} = \lim_{n\to\infty} \frac{\int_A e^{b\pi_n(s)}\,\mathrm{d}p}{\int_B e^{b\pi_n(s)}\,\mathrm{d}p} \\ &= \frac{\lim_{n\to\infty} \int_A e^{b\pi_n(s)}\,\mathrm{d}p}{\lim_{n\to\infty} \int_B e^{b\pi_n(s)}\,\mathrm{d}p} = \frac{\int_A e^{b\pi(s)}\,\mathrm{d}p}{\int_B e^{b\pi(s)}\,\mathrm{d}p}\end{aligned} \tag{33}$$

where the first step follows from Prize-Continuity, the second and fourth since $p_0(B) > 0$ and $p_{\pi_n}(s) \in [m, M]$ on $A \cup B$, the third from Corollary 1, and the fifth from the monotone convergence theorem. □

The final step in the proof of Theorem 1 uses a limit argument whereby a general event A is be approached by events of the form $A_n = \{s \in A : e^{b\pi(s)} \geq 2^{-n}\}$, and Lemma 4 is applied on each of these events separately.

Proof. I prove that if π is a well-behaved distortion then it is a *Priors and Desires* distortion. The opposite direction is trivial. By Lemma 4 there exist a probability measure p_0 and a parameter b such that Equation 32 holds for any events A and B for which $p_0(B) > 0$ and any desires π for which there exists a number $m > 0$ such that $e^{b\pi(s)} \geq m$ on $A \cup B$. To complete the proof I need to show that Equation 32 holds even if no such number m exists. Let π be any desires and let A and B be any events such that $p_0(B) > 0$. If π is almost everywhere simple the claim follows from Corollary 1. Otherwise, by Lemma 3 there exists a number $M \in \mathbb{R}$ such that $e^{bx} \leq M$ for all $x \in X$. For $n \in \mathbb{N}$ let $A_n = \{s \in A : e^{b\pi(s)} \geq 2^{-n}\}$, and similarly define B_n. By construction $\lim_{n\to\infty} A \setminus A_n = \emptyset$ and similarly $\lim_{n\to\infty} B \setminus B_n = \emptyset$. Moreover, since $p_0(B) > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $p_0(B_n) > 0$. The conditions for Lemma 4 therefore hold for A_n and B_n for all $n \geq n_0$. Combining these observations we obtain that

$$\begin{aligned}\frac{p_\pi(A)}{p_\pi(B)} &= \lim_{n\to\infty} \frac{p_\pi(A_n)}{p_\pi(B_n)} = \lim_{n\to\infty} \frac{\int_{A_n} e^{b\pi(s)}\,\mathrm{d}p}{\int_{B_n} e^{b\pi(s)}\,\mathrm{d}p} = \frac{\lim_{n\to\infty} \int_{A_n} e^{b\pi(s)}\,\mathrm{d}p}{\lim_{n\to\infty} \int_{B_n} e^{b\pi(s)}\,\mathrm{d}p} \\ &= \frac{\int_A e^{b\pi(s)}\,\mathrm{d}p}{\int_B e^{b\pi(s)}\,\mathrm{d}p}\end{aligned} \tag{34}$$

where step 3 holds since the integrals are bounded from below and above: (i) $e^{bx} \leq M$ for all $x \in X$, so the integrals are bounded from above by M, and (ii)

$p_0(B_{n_0}) > 0$ and $\pi(s) \geq 2^{-n_0}$ on B_{n_0}, and hence there exists some $m > 0$ such that for all $n \geq n_0$, $\int_{B_n} e^{b\pi(s)}\,\mathrm{d}p \geq \int_{B_{n_0}} e^{b\pi(s)}\,\mathrm{d}p \geq 2^{-n_0} p_0(B_{n_0}) \geq m$. $\square$

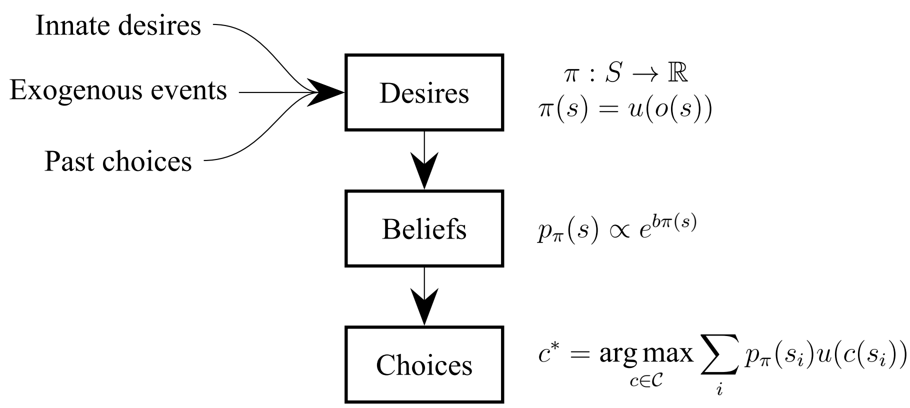

Figure 1: Model outline. What wishful thinkers want to be true affects what they believe to be true and consequently the decisions that they make.

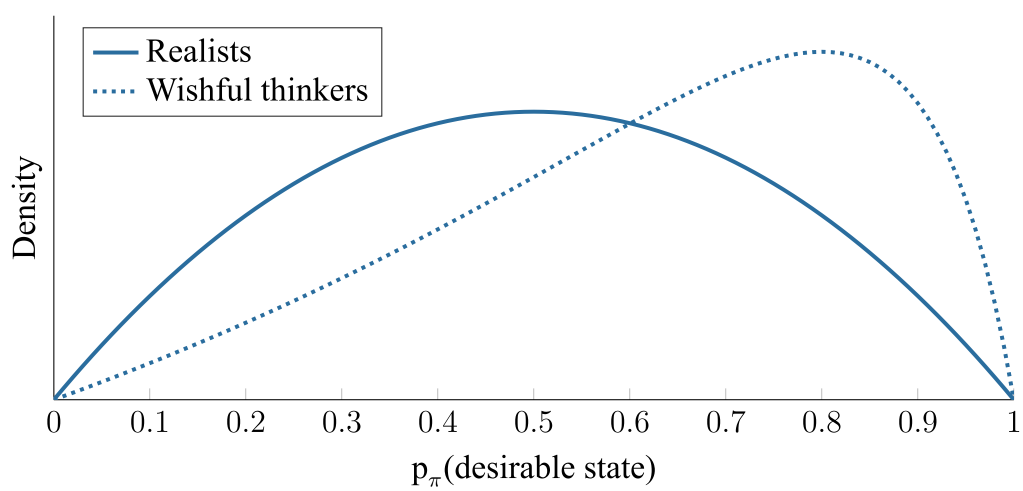

Figure 2: Wishful thinking changes the distribution of beliefs in the population, but as long as the original distribution has full support, individual wishful thinkers can continue to think of themselves as realists. The distribution of realists' beliefs in this chart is modeled by a beta distribution with parameters $\alpha = \beta = 2$. The distribution of wishful thinkers' beliefs is obtained from Equation 1 by assuming that p_0 has the same distribution as realists' beliefs and that $b(\pi(\text{desirable state}) - \pi(\text{other state(s)})) = 4$.