Priors and Desires

Guy Mayraz*

April 6, 2017

Abstract

This paper offers a simple but powerful model of wishful thinking, cognitive dissonance, and related biases. Choices maximize subjective expected utility, but beliefs depend on the decision maker's interests as well as on relevant information. Simplifying assumptions yield a representation in which the payoff in an event affects beliefs as if it were part of the evidence about its likelihood. A single parameter determines both the direction and weight of this 'evidence', with positive values corresponding to optimism and negative values to pessimism. Changes to a person's interests amount to new 'evidence', and can alter beliefs even in the absence of new information. The magnitude of the bias increases with the degree of uncertainty and the strength of the decision maker's interests. High stakes can reduce the bias indirectly by increasing incentives to acquire information, but are otherwise consistent with substantial bias. Exploring applications, I show that wishful thinking can lead investors to become progressively more exposed to risk, and that while improved policing unambiguously deters crime, increased punishment may have little or no deterrent value.

JEL classification: D01,D03,D80,D81,D83,D84.
Keywords: wishful thinking, cognitive dissonance, reference-dependent beliefs, reference-dependent preferences.

*Department of Economics, University of Melbourne. guy.mayraz@unimelb.edu.au.

1 Introduction

Beliefs depend not only on what people *know* to be true, but also on what they *want* to be true. This paper introduces a model of decision making that allows for this possibility, and uses a number of simplifying assumptions to obtain a tractable and generally applicable representation. The model provides a unified account of wishful thinking, overoptimism, overconfidence, cognitive dissonance, and unrealistic pessimism, and can be used to study their implications for decision making. The psychology evidence for these biases spans decades. The economics evidence is much less extensive, but includes a wide variety of situations.[1] Theoretical applications extend into additional areas.[2]

This paper is hardly the first to offer a model of these biases, but it differs markedly from the previous literature. The common approach is to model biased beliefs as an optimal delusion: decision makers start the planning horizon with unbiased beliefs, and choose a distorted prior so as to maximize their total discounted utility, including utility from anticipation Akerlof and Dickens, 1982; Brunnermeier and Parker, 2005.[3]

Optimal delusion models can plausibly explain many cases of optimistic bias. For example, a moderate bias over health risks can be seen as a tradeoff between the desire to minimize fear and the possibility that biased beliefs would result in behavior that would make disease more likely. A moderate level of bias is, however, predicated on relevant choices having correspondingly moderate stakes. In the limit of low stakes the bias is extreme, and in the

[1] Babcock and Loewenstein (1997) find that parties in negotiations are affected by wishful thinking, resulting in an inefficient failure to reach agreement. Camerer and Lovallo (1999) link excess entry into competitive markets to overconfidence over relative ability. Malmendier and Tate (2008) argue that managerial overconfidence is responsible for corporate investment distortions. Cowgill et al. (2009) find optimistic bias in corporate prediction markets. Mullainathan and Washington (2009) find that voting for a candidate results in more positive views about the candidate. Park and Santos-Pinto (2010) provide field evidence for overconfidence in tournaments. Mayraz (2011) finds that a person's expectations for the future price of a financial asset depend on whether he or she gains from high or low prices. Hoffman (2011a), Hoffman (2011b) finds that truck drivers are optimistically biased about their productivity (and hence their pay), resulting in an inefficient failure to switch jobs.

[2] For example, credit markets De Meza and Southey, 1996, banking Manove and Padilla, 1999, corporate finance Heaton, 2002, search Dubra, 2004, savings Brunnermeier and Parker, 2005, insurance Sandroni and Squintani, 2007, price discrimination Eliaz and Spiegler, 2008, incentives in organizations Santos-Pinto, 2008, and financial contracting Landier and Thesmar, 2009. Studies of overconfidence over the accuracy of signals are excluded from this list.

limit of high stakes it disappears entirely. By assumption, it is not possible to model situations in which the bias leads to welfare loss in expectation. Optimism in such models is related to risk and ambiguity preferences. Since riskier bets offer more scope for bias, optimism implies a *ceteris paribus* preference for risk. If optimism is restricted to beliefs over ambiguous bets, it further implies a *ceteris paribus* preference for ambiguity.[4]

This paper's approach is very different. Decision makers take their beliefs as given, and maximize subjective expected utility in their choices. Their beliefs may, however, depend on what they want to be true *prior* to making their choice. A number of simplifying assumptions are imposed, yielding a representation with one non-standard parameter, which determines both the direction and magnitude of the bias.

Some predictions are broadly similar to those of optimal delusion models. For example, optimists underestimate their health risks, much as they would in an optimal delusion model. The degree of bias is, however, independent of the impact on future choices, and a substantial bias may occur even if its cost is high. Decision makers are only biased about a choice if its outcome hangs on events that form part of their existing interests. There is thus no general tendency to favor risky or ambiguous bets. On the other hand, choices that align (conflict) with an optimist's existing interests are perceived to be relatively likely (unlikely) to lead to desirable outcomes. This feature of the model leads to path dependence: an optimist who is invested in some asset is biased in favor of increasing her exposure to the asset and against replacing her existing investment with an opposite bet.[5]

Changes to the decision maker's interests affect beliefs, even if the relevant information is unchanged. This makes it possible to model cognitive dissonance as a dynamic version of wishful thinking. For example, Knox and Inkster (1968) find that when bettors commit to place a bet on a horse, their confidence that the horse would win the race goes up. This finding can be read-

[3]There are a number of papers in which agents manipulate their belief indirectly, by strategically choosing what information to consume Carrillo and Mariotti, 2000; Kőszegi, 2006 or by a biased memory process Benabou and Tirole, 2002; Compte and Postlewaite, 2004; Gottlieb, 2010.

[4]The connection between optimism (or pessimism) and preferences for risk and ambiguity is also shared by such models as Hey (1984), Bracha and Brown (2010) and Dillenberger et al. (2011), in which the probabilities used to evaluate an alternative vary with the payoffs in that alternative.

[5]In a labor context, such path dependence would lead to an inefficiently low rate of quitting, consistent with the findings in Hoffman (2011a), Hoffman (2011b).

ily explained by the change in interests: the bettors are initially indifferent as to which horse would win the race, but once they place the bet they gain an interest in 'their' horse winning. Beliefs are thus initially unbiased, but become biased once the bet is placed. Exogenous changes in interests are also predicted to affect beliefs. For example, the finding that congressmen become more positive about women's interests after fathering a daughter Washington, 2008 can be explained by the consequent increase in alignment between a congressman's interests and those of women.[6]

The core of the model is the precise relationship between a decision maker's beliefs and her interests. Subjective beliefs are represented by a probability measure over the set of states, and interests are represented by a *payoff-function*, or the mapping associating each state with the utility that the decision maker obtains in that state. Letting f denote the payoff-function, I let π_f denote the resulting probability measure. In order to obtain a tractable representation, I make a number of simplifying assumptions, which take the form of special circumstances in which different payoff-functions do *not* result in different beliefs.

The formal framework and simplifying assumptions are presented in Section 2. A representation theorem establishes that the assumptions are necessary and sufficient conditions for the existence of a probability measure p and a real-valued parameter ψ, such that for any payoff-function f and any event A,

$$\pi_f(A) \propto \int_A e^{\psi f} \, \mathrm{d}p. \tag{1}$$

Equation 1 takes a simpler form if the state-space is discrete, when it can be written as follows, where s is any state:

$$\pi_f(s) \propto p(s) e^{\psi f(s)}. \tag{2}$$

To understand this representation note first that if f is constant then $\pi_f = p$. The probability measure p therefore represents the decision maker's *indifference beliefs*, or the beliefs she would hold if she were a completely disinterested observer who is indifferent between all states. More generally, f is not constant, and π_f depends on both f and on ψ. If ψ is positive (negative) π_f is higher in states in which payoff is higher (lower). A positive value of ψ therefore represents *optimistic* bias, and a negative value represents *pessimistic* bias.

[6]Socialization is an alternative explanation Washington, 2008.

The larger ψ is in absolute terms, the greater the bias. In analogy with relative risk-aversion, ψ can be thought of as the *coefficient of relative optimism*. Standard decision makers are represented by $\psi = 0$. Such decision makers are *realists*, and for them $\pi_f = p$ for all f.

Section 3 completes the model, and provides it with a revealed preferences axiomatization. The overall framework is that of reference-dependent preferences, where preferences are over alternatives (acts), and the reference corresponds to the mapping from states to consequences that characterizes the decision maker prior to the new choice. Axioms ensure (i) that holding the reference constant, preferences have a subjective expected utility representation, (ii) that the utility function is *not* reference-dependent (making this a model of reference-dependent *beliefs*), (iii) that the reference affects beliefs only via the interests that follows from it, and (iv) that beliefs relate to interests via Equation 1.

In the representation that follows from the axioms, each decision maker is defined by a utility function u, a probability measure p, and a coefficient of relative optimism ψ.[7] Let f denote the payoff-function representing the decision maker's interests before being presented with the choice set. According to the model, her choice maximizes the expectation of u given π_f, where π_f is given by Equation 1.

One interpretation of the model is that subjective judgment is in some people biased in an optimistic directions, and in others in a pessimistic direction. Decision makers have standard tastes, and seek to maximize their (true) expected utility. Frequently, however, they find themselves having to resort to subjective judgment when assessing the likelihood of events,[8] and in those cases they end up with biased beliefs. Biased beliefs then lead to biased choices.

The model with $\psi > 0$ provides a unified account of wishful thinking and cognitive dissonance. These two biases can be seen, respectively, as simply the cross-sectional and time-series manifestations of optimism. Overopti-

[7] ψ and u are identified together, and the representation is unique up to a positive affine transformation. If u is replaced by $u' = au + b$, ψ must be replaced by $\psi' = \psi/a$.

[8] Knight (1921) emphasized the importance of subjective judgment in decisions under uncertainty: "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." (III.VII.47); "Yet it is true, and the fact can hardly be overemphasized, that a judgment of probability is actually made in such cases." (III.VII.40).

mism and overconfidence in ability can be viewed as special cases of wishful-thinking.[9] The model with $\psi < 0$ can be used to represent unrealistic pessimism Seligman, 1998.

Section 4 takes a closer look at the equations relating beliefs to interests. An important insight is that they are formally identical to Bayes Rule, with p standing for prior beliefs, π_f for posterior beliefs, and $\psi f(s)$ for the log likelihood in state s. It is thus as if optimists (pessimists) are Bayesian belief updaters, who believe that Nature has chosen the state of the world so as to make life better (worse) for them.

The comparative statics of the model are that the bias is increasing in the degree of optimism or pessimism, the strength of the decision maker's interests, and the degree of subjective uncertainty.[10] High-stakes decisions result in smaller bias only if the increased incentive to invest in information reduces the decision maker's uncertainty. If good information is not available, a substantial bias may remain in spite of its costly consequences.[11]

Section 5 examines belief updating. Changes to the decision maker's interests can alter beliefs even in the absence of new information, as in the Knox and Inkster (1968) study of horse bettors. A more subtle phenomenon appears whenever the decision maker's interests involve a non-linear function of events. News about one event can then alter the subjective probability of the other, even if the two events are independent. For example, a professor's belief about the importance of publishing in a top journal may depend on whether her paper is accepted.

When using the model in a new application it is necessary to introduce modeling assumptions for p and for ψ. One option is to reinterpret rational expectations as applying to the indifference beliefs p. Optimistic bias can be modeled simply by assuming that ψ is positive. Consider the beliefs of investors during an asset bubble. The implication of rational expectations is that investors with no exposure to the asset hold unbiased beliefs about the prospect

[9] Overoptimism is exemplified by the Weinstein (1980) finding that students believe desirable (undesirable) life events are more (less) likely to happen to them than to other students. The Svenson (1981) finding that most people believe themselves to be better drivers than most other people is the best known example of overconfidence.

[10] Intuitively, stronger optimism/pessimism and stronger interests correspond to stronger 'evidence', whereas more subjective uncertainty implies a weaker 'prior'.

[11] The magnitude of the bias is measured by its effect on the subjective odds-ratio between events. Since certainty corresponds to an infinite odds-ratio, no amount of bias can cause a certain event to be perceived as less than certain (or the other way around.)

of market collapse. The assumption that $\psi > 0$ implies that investors who hold the asset underestimate this probability.

The model makes it possible to adapt existing applications to incorporate the implications of wishful thinking and cognitive dissonance. Suppose an agent in some model maximizes expected utility given beliefs p and utility function u. In adapting the model to allow for wishful thinking, the natural assumption is that p represents the agent's indifference beliefs. Equation 1 can be used to compute the distorted probability measure π_f, which can then be used in place of p in predicting the agent's choices.

Section 6 presents two applications. The first shows that an optimism leads investors to escalate risky investments. The intuition is that taking up a risky investment creates an interest in the risk being low, and the investor consequently comes to believe the risk is lower than it really is. Given the revised risk assessment, the investor feels secure in increasing the investment. The second application is to the economics of crime. There is much evidence that increasing the *severity* of punishment is a relatively ineffective deterrent as compared with increasing the *likelihood* of punishment Grogger, 1991; Nagin and Pogarsky, 2001; Durlauf and Nagin, 2011. The model can explain this finding on the assumption that criminals are optimistic. The intuition is that increasing the punishment gives criminals a stronger interest in not getting caught, resulting in a bigger bias in their beliefs. Thus, while getting caught is *worse*, it is also subjectively *less likely*. By contrast, increasing the likelihood that criminals are brought to justice leaves the bias in their beliefs unchanged, and unambiguously improves deterrence.

The model stands in an interesting relationship to models of reference-dependent utility, such as Kőszegi and Rabin (2006). In both types of model preferences depend on a reference act, but in Kőszegi and Rabin (2006) the *utility* of different outcomes is dependent on the *probability* in which these outcomes are obtained, while in this model the *probability* of different states is dependent on the *utility* in those states. In Kőszegi and Rabin (2006) it makes no difference in which particular states a given consumption outcome is obtained (only the overall probability matters). In this model it makes no difference which particular outcome is obtained in a given state (only its utility matters).

2 Belief distortion

The core of the model is the relationship between people's beliefs and what they want to be true. In this section I state and prove a representation theorem that characterizes this relationship. Section 2.1 describes the framework that links beliefs and interests, the properties that are assumed to characterize this link, and the formal statement of the theorem. Section 2.2 demonstrates the role of the individual assumptions, by presenting the partial representation results that can be obtained with only a subset of the assumptions. The proof is described for the special case in which there are only finitely many events, making it possible to focus on the key ideas, while avoiding the technical complications that arise in the more general case. Section 2.3 concludes the proof of the representation theorem by extending this result to any measurable space.

2.1 Framework

Subjective uncertainty is defined over a measurable-space (S, Σ), where S is the set of *states*, and Σ is a σ-algebra of subsets of S, called *events*. The decision maker's interests are represented by a *payoff-function*, which is a mapping associating each state with the utility that is obtained in that state. I let $X = [m, M] \subseteq \mathbb{R}$ denote the set of all feasible payoffs, which I assume to be an interval which includes 0. A payoff-function is formally a Σ-measurable mapping $f : S \to X$. Let F denote the set of all such functions, and let Δ denote the set of all σ-additive probability measures over (S, Σ). The key ingredient in the model is a *distortion mapping* $\pi : F \to \Delta$, associating with each payoff-function a probability measure over (S, Σ). The distortion mapping π is the formal representation of the possibility that a person's beliefs are a function of her interests. The goal of this section is to develop a tractable representation for π.

In the following definitions f and f' stand for any payoff-functions, a for any constant payoff-function, and E for any event. The first definition states the properties we want the distortion mapping to satisfy, and the second describes the logit-distortion formula. The theorem says that the two definitions are equivalent.

Definition 1. $\pi : F \to \Delta$ is a *well-behaved distortion* if the following conditions are satisfied:

A1 (absolute continuity) $\pi_{f'}(E) = 0 \iff \pi_f(E) = 0$.

A2 (consequentialism) If $f = f'$ over a non-null[12] event E then $\pi_{f'}(\cdot|E) = \pi_f(\cdot|E)$.

A3 (shift-invariance) If $f' = f + a$ then $\pi_{f'} = \pi_f$.

A4 (prize-continuity) If $f_n \to f$ then $\pi_{f_n}(E) \to \pi_f(E)$.

These properties should be understood as simplifying assumptions, the purpose of which is to obtain as simple as possible a representation, while retaining the ability to represent the phenomena we wish to model. *Absolute Continuity* limits belief distortion to events that the decision maker is uncertain about. *Consequentialism* requires that if two payoff-functions coincide over some event E then the corresponding probability measures *conditional* on E also coincide. Consider two events E_1 and E_2 that are subsets of E and an event F that is outside E. According to Consequentialism, a change in the payoff in F can affect the overall probability of E_1 and E_2, but it cannot affect their relative probability. *Shift Invariance* requires subjective probabilities to depend only on payoff *differences* between states. A person's interests in an event being true are defined by how much she has to gain or lose (in utility terms) if the event is true. Increasing all payoffs by a constant leaves interests unchanged, and should not result in a change to beliefs. *Prize Continuity* requires that small differences in payoffs have only a small effect on beliefs.

Definition 2 (Logit distortion)**.** $\pi : F \to \Delta$ is a *logit distortion* if there exists a probability measure p (the *indifference measure*), and a real-number ψ (the *coefficient of relative optimism*), such that for any payoff-function f and any event A,

$$\pi_f(A) \propto \int_A e^{\psi f} \, \mathrm{d}p. \tag{3}$$

Consequentialism only has bite when there are at least three events with positive probability. This condition is therefore necessary for the equivalence between the two definitions to hold.

Definition 3 (Minimally complex distortion)**.** $\pi : F \to \Delta$ is *minimally complex* if there exists three disjoint events A, B, and C, and a payoff-function f such that $\pi_f(A)$, $\pi_f(B)$, and $\pi_f(C)$ are all positive.

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

[12]That is, both $\pi_f(E) > 0$ and $\pi_{f'}(E) > 0$. Absolute Continuity ensures that these two requirements coincide.

2.2 Intermediate representation results

In this section I prove the theorem for the special case where there are only 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 $\mathcal{S}$. In addition, I prove a sequence of partial representation results requiring only a subset of the assumptions. In order to state the necessary and sufficient conditions for these representations I define a new property, *Indifference*, which is related to Shift Invariance, but is considerably weaker:

A3' (Indifference). $\pi_f = \pi_{f'}$ if both f and f' are constant payoff-functions.

Note that unlike Shift Invariance, Indifference does not require the set of payoffs to have cardinal (or even ordinal) meaning.

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 $p \in \Delta$ and a function $h : F \times \mathcal{S} \to \mathbb{R}_+$, such that for any payoff-function f and any event $A \in \mathcal{S}$,*

$$\pi_f(A) \propto p(A)\ h_f(A). \tag{4}$$

2. *Assume Absolute Continuity. Consequentialism is a necessary and sufficient condition for there to exist a probability distribution $p \in \Delta$, and a mapping $\mu : \mathcal{S} \times X \to \mathbb{R}_+$, such that for any payoff-function f and any event $A \in \mathcal{S}$,*

$$\pi_f(A) \propto p(A)\ \mu_A(f(A)). \tag{5}$$

3. *Assume Absolute Continuity and Consequentialism. Indifference is a necessary and sufficient condition for there to exist a probability distribution $p \in \Delta$, and a mapping $\nu : X \to \mathbb{R}_+$, such that for any payoff-function f and any event $A \in \mathcal{S}$,*

$$\pi_f(A) \propto p(A)\ \nu(f(A)). \tag{6}$$

4. *Assume Absolute Continuity and Consequentialism. Shift-Invariance and Prize-Continuity are necessary and sufficient conditions for there*

to exists a probability distribution $p \in \Delta$, and a parameter $\psi \in \mathbb{R}$, such that for any payoff-function f and any event $A \in \mathcal{S}$,

$$\pi_f(A) \propto p(A)\, e^{\psi f(A)}. \tag{7}$$

Note that while the representation in Equations 4–7 is defined with respect to events in $\mathcal{S}$, the implication for general events is straightforward.[13] The following simple example demonstrates that Minimal Complexity is a necessary assumption. Let $\mathcal{S} = \{A, B\}$, let $\pi_f(A) \propto p(A)(1 + (f(A) - f(B))^2)$ and $\pi_f(B) \propto p(B)$. This distortion is well-behaved (Definition 1), but it cannot even be given the representation of Equation 5, let alone that of a logit distortion (Definition 2).

2.3 Completing the proof

This section concludes the proof of Theorem 1 for the general case. The first step is to generalize Equation 7 to any payoff-function and any constant-payoff events:

Lemma 2. *Suppose $\pi : F \to \Delta$ is a minimally complex well-behaved distortion, then there exist a probability measure p and a parameter $\psi \in \mathbb{R}$, such that for any payoff-function f and any events A and B such that $p(B) > 0$ and f is constant on A and on B,*

$$\frac{\pi_f(A)}{\pi_f(B)} = \frac{p(A)}{p(B)}\, \frac{e^{\psi f(A)}}{e^{\psi f(B)}}. \tag{8}$$

Theorem 1 for simple payoff-functions is an immediate corollary.[14] The following claim is a little more general, allowing for functions that are almost everywhere simple:

Definition 4. A payoff-function $f \in F$ is *almost everywhere simple* if there exists a payoff-function $g \in F$ and an event E such that f obtains only finitely many values on E and $\pi_g(E) = 1$.

Corollary 1. *Theorem 1 holds when restricted to payoff-functions that are almost everywhere simple.*

[13] Any event in Σ is the finite union of events in $\mathcal{S}$.

[14] A payoff-function f is simple if $f(S)$ is finite.

The remaining case involves functions which are *not* almost everywhere simple. If such payoff-functions exist, there must also exist an infinite sequence of non-null events $\{A_n\}_{n\in\mathbb{N}}$. But then, as long as $\psi \neq 0$ and the set of feasible payoffs is unbounded, it is possible to construct a payoff function f such that $\lim_{n\to\infty} \frac{\pi_f(A_n)}{\pi_f(A_1)} = \infty$. But this implies that $\pi_f(A_1) = 0$, in contradiction to Absolute Continuity. Hence, if $\psi \neq 0$ the set of feasible payoffs must be bounded.

Lemma 3. *Suppose $\pi : F \to \Delta$ is a minimally complex well-behaved distortion, and that there exists a payoff-function f that is not everywhere simple, then there exist an upper bound $M \in \mathbb{R}$, such that for any feasible payoff-value x, $e^{\psi x} \leq M$.*

Lemma 3 ensures that $e^{\psi X}$ is bounded from above. If it is also bounded from below, a limit argument based on simple payoff-functions can be used to extend the claim further:

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

$$\frac{\pi_f(A)}{\pi_f(B)} = \frac{\int_A e^{\psi f}\, dp}{\int_B e^{\psi f}\, dp}. \tag{9}$$

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

3 Preferences

Section 2 characterizes the relationship between beliefs and interests. This section completes the model, by embedding this dependence in a complete model of choice, and providing it with a revealed-preferences axiomatization. The overall framework is that of reference-dependent preferences, as is the case in models of reference-dependent utility. The axioms in this section, however, ensure that this model is one of reference-dependent *beliefs*. Further axioms ensure that the particular consequences that are obtained in different states are irrelevant, and that the only thing that matters is the utility value associated with any given consequence. In other words, a reference act affects beliefs only via the associated payoff-function. A final set of axioms restates the simplifying assumptions of Section 2 in revealed preferences terms. The resulting representation comprises three elements that are determined together: a utility-function, a probability-measure, and a real-valued coefficient of relative-optimism.

Both the reference and the choices are *acts*, or mappings from states to consequences. For example, in an investment application, the reference may be an investor's current portfolio, represented by a mapping from market outcomes to different amounts of money, and the choice set may be a selection of alternative portfolios. More generally, consequences can also be objective lotteries over final outcomes, making it possible to model situations in which a person takes some probabilities as given. A bet on a coin-toss, for example, involves no subjective uncertainty, and is represented by an act mapping all subjective states to the same 50-50 lottery over the possible outcomes of the bet. More interestingly, it is possible to model uncertainty over which of several probabilistic models is correct. For example, a smoker may be uncertain whether smoking increases her risk of getting cancer. States correspond to whether this is or isn't the case, and the smoker's situation is described by an act mapping these states into different probabilities of cancer. A non smoker's situation would be described by an act mapping both states to the same low probability.[15]

3.1 Framework

Let S denote the set of subjective states, and let Σ be a σ-algebra of events. I let Z denote the set of *final outcomes*. An (Anscombe-Aumann) *act* is a Σ-measurable mapping $a : S \to \Delta(Z)$, associating with each subjective state an objective lottery over final outcomes.[16] I denote by $\mathscr{A}$ the set of all acts. The key object is a set of reference-dependent preferences $\succeq: \mathscr{A} \to 2^{\mathscr{A}}$, associating with each act $e \in \mathscr{A}$ a preference relation $\succeq_e$. The interpretation of $g \succeq_e h$ is that the decision maker prefers g over h *if her reference act is e*.

[15] Assuming both are optimistic, the smoker—but not the non-smoker—is therefore likely to underestimate the link between smoking and cancer.

[16] Acts mapping states into objective lotteries over final consequences were introduced by Anscombe and Aumann (1963).

In the following let g, h, e, e' and e_n denote general acts, and let a and b denote constant acts.[17] Let s and s' denote general states, and let E denote a general event. For an act e and a state s, let e_s denote the constant act yielding $e(s)$ in all states. Let $e_s \sim e_{s'}$ if for all acts d, $e_s \sim_d e'_s$. Finally, an event E is $\succeq_e$ *null* if $g \sim_e h$ for all g and h that differ only on E. The assumptions on $\succeq$ are as follows:

B1 (*Anscombe-Aumann*) For all $e \in \mathscr{A}$, $\succeq_e$ has an Anscombe-Aumann expected utility representation.

B2 (*objectivity*) For all $e, e' \in \mathscr{A}$, $\succeq_e = \succeq_{e'}$ over constant acts.

B3 (*indifference*) If $e_s \sim e'_s$ for all s then $\succeq_e = \succeq_{e'}$.[18]

B4 (*non-triviality*) For any act e there exist constant acts a and b such that $a \succ_e b$.

B5 (*best and worst act*) For any act e there exist constant acts $\overline{a}$ and $\underline{a}$ such that for any act g, $\overline{a} \succeq_e g \succeq_e \underline{a}$.

B6 (*absolute continuity*) E is $\succeq_e$ null $\iff$ E is $\succeq_{e'}$ null.

B7 (*consequentialism*) If $e = e'$ over E and $g = h$ outside E then $g \succeq_e h \iff g \succeq_{e'} h$.

B8 (*shift-invariance*) If for some $\alpha \in [0,1]$, $e = \alpha g + (1-\alpha)a$ and $e' = \alpha g + (1-\alpha)b$, then $\succeq_e = \succeq_{e'}$.[19]

B9 (*continuity*) If $e_n \to e$ uniformly[20] then $\succ_{e_n} \to \succ_e$.[21]

[17] A constant Anscombe-Aumann act yields the same objective lottery in all states.

[18] That is, for e and e' to result in different preferences, it is not enough that $e(s) \neq e'(s)$ for some state s—it is also necessary that one of these outcomes is strictly preferred to the other (formally, the decision maker has strict preferences between the constant acts e_s and e'_s. B2 ensures that these preferences are well-defined.)

[19] This condition implies that the utility difference between e and e' is the same in all states.

[20] For any $\epsilon > 0$ and any state $s \in S$ there exists $n_0 \in \mathbb{N}$, such that for all $n > n_0$ and for any outcome $z \in Z$, $|e_n(s)(z) - e(s)(z)| < \epsilon$ (the difference in the probability the two acts assign to outcome z in state s is less than ϵ)

[21] For all acts g and h, if $g \succ_e h$ then there exists $n_0 \in \mathbb{N}$, such that for all $n > n_0$, $g \succ_{e_n} h$.

Let $h \in \mathscr{A}$ be any act mapping states to lotteries over final outcomes, and let $u : Z \to \mathbb{R}$ denote a function mapping final outcomes to real numbers. I use the notation $uh : S \to \mathbb{R}$ to denote the mapping from states to real numbers that is obtained by taking the expected value of u in each state. That is, $uh(s) = \int_Z u(h_s(z))\,\mathrm{d}z$.

The representation we wish to obtain is the following:

Definition 5. The reference-dependent preferences $\succeq : \mathscr{A} \to 2^{\mathscr{A}}$ are *logit preferences* if there exist a probability measure p over (S, Σ) (the indifference measure), a function $u : Z \to \mathbb{R}$ (the utility function), and a real number ψ (the coefficient of relative optimism), such that for any reference act $e \in \mathscr{A}$, $\succeq_e$ ranks acts according to the following Anscombe-Aumann expected utility functional:

$$V_e(g) = \int_S (ug)\,\mathrm{d}\pi_{(ue)}, \tag{10}$$

where $g \in A$ is any act, $F = \mathrm{f}ue : e \in \mathscr{A}\mathrm{g}$, and $\pi : F \to \Delta(S)$ is a logit distortion (Definition 2). The trio (p, u, ψ) is then said to *represent* $\succeq$.

As in Section 2 (and for the same reason) the proof requires the existence of at least three disjoint non-null events:

Definition 6. $\succeq$ is *minimally complex* if there exist an act e and disjoint events A, B, and C that are not null with respect to $\succeq_e$.

Theorem 2. *Suppose $\succeq$ are reference-dependent preferences and that assumptions B1-B9 hold, then $\succeq$ are logit preferences. Moreover, if both (p, u, ψ) and (p', u', ψ') represent $\succeq$, then $p' = p$ and there exist a positive real-number α and a real-number β, such that $u' = \alpha u + \beta$ and $\psi' = \frac{\psi}{\alpha}$.*

3.2 Proof

B1 is an omnibus axiom, requiring that—conditional on the reference act—preferences have a subjective expected utility representation. Thus, for any reference act e there exists a probability measure $\mu_e \in \Delta(S)$ and a utility function $u_e : Z \to \mathbb{R}$, such that $\succeq_e$ ranks acts according to $V_e(g) = \int_S (u_e g)\,\mathrm{d}\mu_e$. This representation allows for both beliefs and tastes to vary with the reference act e. B2 rules out the latter possibility by imposing the requirement that the ranking of constant Anscombe-Aumann acts does not depend on e. Since the ranking of constant Anscombe-Aumann acts identifies the utility function up at a positive affine transformation, there exists a utility-function u, such that $u_e = u$ for all e. Given that $\succ_e$ ranks acts in accordance with $V_e(g) = \int_S (ug)\,\mathrm{d}\mu_e$, B3 implies that $\mu_e = \mu_{e'}$ whenever $ue = ue'$. Beliefs may, therefore, depend on the reference act only via the associated payoff-function. Let $F = \{ue : e \in \mathscr{A}\}$. Given B3, there exists a mapping $\pi : F \to \Delta(S)$, such that $\mu_e = \pi_{ue}$ for all e.[22] B4 is a technical assumption ruling out the trivial case in which the decision maker is indifferent between all acts. Non-triviality ensures that it is possible to back out π_{ue} from observing $\succeq_e$. Hence, $\succeq_e = \succeq_{e'}$ if and only if $\pi_{ue} = \pi_{ue'}$. We thus obtain the following Lemma:

Lemma 5. *Suppose B1-B4, then (i) there exists a utility function $u : \Delta(Z) \to \mathbb{R}$, and a mapping $\pi : F \to \Delta(S)$ where $F = \{ue : e \in \mathscr{A}\}$, such that for any $e \in \mathscr{A}$, $\succeq_e$ ranks acts in accordance with the following subjective expected utility functional:*

$$V_e(g) = \int_S (ug)\, d\pi_{ue} \tag{11}$$

and (ii) for any two acts e and e', $\succeq_e = \succeq_{e'}$ if and only if $\pi_{ue} = \pi_{ue'}$.

B5 is a second technical assumption, ensuring that there exist a best and a worst lottery (and therefore also a best and a worst outcome). B6-B9 effectively restate assumptions A1-A4 of Definition 1. The proof of the following Lemma is in Appendix A.

Lemma 6. *Suppose B5-B9 hold in addition to B1-B4 then the mapping π in Lemma 5 is a well-behaved distortion.*

The main claim in Theorem 2, namely the existence of a triplet (p, u, ψ) representing $\succeq$ in accordance with Equation 10 and Definition 2, is an immediate corollary of Lemmas 5 and 6 together with Theorem 1. The proof of the uniqueness part is in Appendix A.

4 The belief distortion function

This section takes a close look at the belief distortion function that was derived in Section 2. Let π denote the mapping from payoff-functions to beliefs.

[22] π is formally defined by choosing for any payoff-function f some particular act $e(f)$ to represent the equivalence class of all the acts having f as their payoff-function, and defining $\pi(f) = \mu_{e(f)}$.

According to Theorem 1, if the simplifying assumptions hold, there exists a probability measure p, and a real-valued parameter ψ, such that for any payoff-function f, and any event A,

$$\pi_f(A) \propto \int_A e^{\psi f}\, \mathrm{d}p. \tag{12}$$

Consider first the case where f is constant, representing a situation in which the decision maker is equally well-off in all states, and hence indifferent as to what the true state is. The $e^{\psi f}$ term drops out, and we obtain that $\pi_f = p$. The probability measure p can therefore be identified with a decision maker's *indifference beliefs*, or the beliefs she would hold if she were a disinterested observer. More generally, $e^{\psi f}$ is increasing in the payoff if ψ is positive, decreasing in the payoff if ψ is negative, and independent of it if $\psi = 0$. A positive value of ψ therefore represents *optimistic bias*, a negative represents *pessimistic bias*, and a zero value represents *realism*. The magnitude of belief distortion increases when moving away from zero, whether in the optimistic or pessimistic direction. In analogy with the coefficient of relative risk aversion, ψ is the *coefficient of relative optimism*.

Equation 12 allows for payoffs to vary arbitrarily between different states. If we restrict attention to events over which the payoff is constant, we can rewrite the equation as follows:

$$\pi_f(A) \propto p(A)\, e^{\psi f(A)}. \tag{13}$$

Further insight can be obtained by comparing the probability of two events in relation to each other. Suppose that f is constant over two events A and B, and that B is not-null.[23] The log-odds ratio between the two events can be written as follows:

$$\log \frac{\pi_f(A)}{\pi_f(B)} = \log \frac{p(A)}{p(B)} + \psi\, \left[f(A) - f(B)\right]. \tag{14}$$

The bias in the relative probability of two events depends only on the payoff-difference, or the degree to which one is more desirable than the other. If a decision maker is indifferent between two events, their relative probability is unchanged.

More generally, the same subjective probabilities may result from different combinations of *interests* (represented by the payoff-difference $f(A) - f(B)$)

[23] By Absolute Continuity, the set of non-null events is the same for all reference payoff-functions.

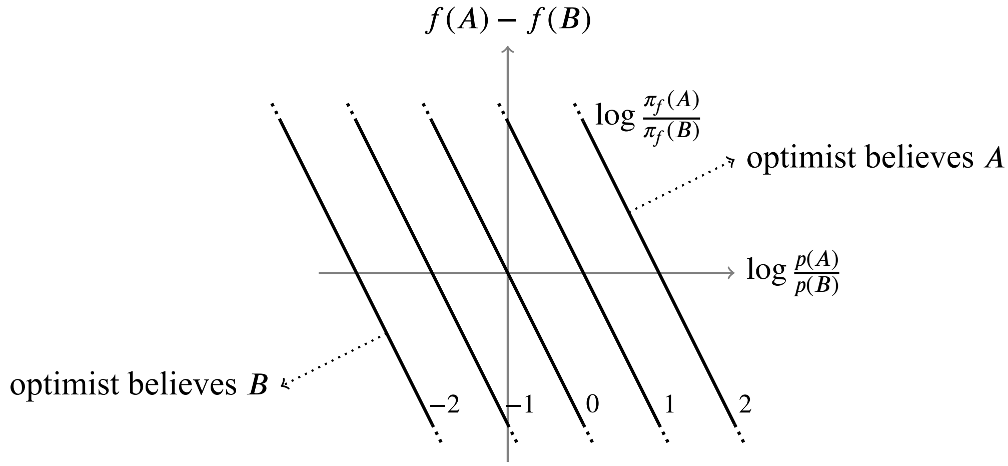

Figure 1: Iso-belief lines for an optimist as a function of the undistorted log odds-ratio on the x-axis and the payoff difference between the two events on the y-axis. Iso-belief lines are straight and slope down and to the right with a slope of $\frac{1}{\psi}$. The Iso-belief lines for pessimists slope upward and to the right. Those of a realist are vertical.

and *information* (represented by the indifference log odds-ratio $\log \frac{p(A)}{p(B)}$). Since Equation 14 is linear, the resulting *iso-belief* lines are also linear (Figure 1).

4.1 Payoffs as information

The equations of the model have a close analogue in Bayes Rule. For Equation 13 the analogous equation is the following:

$$p(A|e) \propto p(A)\ L(e|A), \tag{15}$$

where e represents new evidence, p represents beliefs prior to observing the new evidence, $p(A|e)$ represents posterior beliefs, and $L(e|A)$ the likelihood of the new evidence. Similarly, the analogue of Equation 14 is

$$\log \frac{p(A|e)}{p(B|e)} = \log \frac{p(A)}{p(B)} + \log \frac{L(e|A)}{L(e|B)}, \tag{16}$$

where $\frac{p(A)}{p(B)}$ is the prior odds ratio, $\frac{p(A|e)}{p(B|e)}$ is the posterior odds ratio, and $\frac{L(e|A)}{L(e|B)}$ is the likelihood ratio. A comparison of these equations reveals a perfect correspondence, with p standing for *indifference* or *prior* beliefs, π_f for *distorted* or *posterior* beliefs, and $\psi f(A)$ for the log-likelihood of the evidence in A, with an analogous expression for B.

It is thus possible to view optimism and pessimism as a Bayesian update on an expanded state-space. Starting with indifference beliefs (represented by p) as her prior, the decision maker observes the payoff-function f, and updates her beliefs to arrive at the posterior π_f. The payoff in an event functions as evidence about its likelihood: an optimist (pessimist) takes high payoff to be evidence that an event is more (less) likely. It is as if optimists (pessimists) believe that nature is not an indifferent force, but is instead well-disposed (ill-disposed) toward them. Given that nature took their interests into account when choosing the state, the can make inferences about nature's choice by observing what their interest are.[24]

4.2 Optimism and pessimism

Suppose a decision maker's beliefs are reference dependent with some distortion mapping π, and that some particular probability measure p represents her beliefs whenever she is indifferent between all states. Let $P_f(x) = p(f \leq x)$ denote the indifference cumulative distribution function (CDF) of payoff, and let $\Pi_f(x) = \pi_f(f \leq x)$ denote the corresponding CDF for π_f. For two distributions F and G let $F \succeq_1 G$ if F first-order stochastically dominates G, and $F \succeq_{\text{LR}} G$ if F stochastically dominates G in the likelihood ratio.[25] If we identify a better payoff distribution with first-order stochastic dominance, we can give optimism and pessimism the following definition:

Definition 7. A decision maker is an *optimist* (*pessimist*) if $\Pi_f \succeq_1 P_f$ ($P_f \succeq_1 \Pi_f$). A decision maker who is both an optimist and a pessimist is a *realist*.

The following proposition establishes the relationship between this definition and the coefficient of relative optimism ψ:

Proposition 1. *Suppose a decision maker's beliefs are characterized by a logit distortion with a coefficient of relative optimism ψ, then the decision maker is an optimist (pessimist) if and only if $\psi \geq 0$ ($\psi \leq 0$). Moreover, $\psi \geq 0 \Rightarrow \Pi_f \succeq_{LR} P_f$ and $\psi \leq 0 \Rightarrow P_f \succeq_{LR} \Pi_f$.*

[24] Compare the 'pessimistic' interpretation of certain models of ambiguity aversion, where a malevolent nature chooses the state of the world *after* after the decision maker makes a choice.

[25] That is, if there exists a non-decreasing function $h : \mathbb{R} \to \mathbb{R}_+$, such that $F(x) \propto \int_{-\infty}^{x} h(x) dG(x)$.

Logit distortions are therefore a tractable subset of optimistic and pessimistic beliefs, much as the class of constant relative risk-aversion (CRRA) preferences is a tractable subset of risk seeking and risk averse preferences.

The higher (lower) ψ is, the more probability shifts toward the states with the highest (lowest) possible payoff. If there are only finitely many payoff values, the limit is always well-defined, and takes a particularly simple form: an extreme optimist (pessimist) is certain she would obtain the best (worst) possible payoff:

Proposition 2 (extreme optimism/pessimism)**.** *Let f be a simple payoff-function, and let A_{min} and A_{max} denote respectively the event that the minimal (maximal) payoff is obtained, then* $\lim_{\psi\to\infty} \pi_f(A_{max}) = \lim_{\psi\to-\infty} \pi_f(A_{min}) = 1$.

One case of particular interest is when payoff is linear in some normally distributed random variable. When that is the case, optimism and pessimism simply shift the mean of the distribution, the shift being proportional to the variance and to the coefficient of relative optimism:

Proposition 3 (normally distributed payoffs)**.** *Suppose $X : S \to \mathbb{R}$ is a random variable with indifference distribution $P_X \sim \mathcal{N}(\mu, \sigma^2)$, and that there exist $a, b \in \mathbb{R}$, such that the payoff-function is $f = aX + b$, then $\Pi_X \sim \mathcal{N}(\mu + \psi a\sigma^2, \sigma^2)$.*

4.3 Comparative statics

The intuition for the comparative statics can be obtained by writing Equation 14 qualitatively as follows:

$$\text{beliefs} = \text{indifference beliefs} + \psi \text{ interests}. \tag{17}$$

The magnitude of the bias is thus increasing in the strength of the decision maker's interests and decreasing in the sharpness of indifference beliefs (increasing in the degree of uncertainty). This is seen most clearly if payoff is linear in a normally distributed random variable, i.e. $f(s) = aX(s) + b$, where $X \sim \mathcal{N}(\mu, \sigma^2)$. According to Proposition 3 the distorted probability density function is also normal, the variance is the same, and the mean is shifted in proportion to $\psi a\sigma^2$. The bias is thus increasing in the strength of interests a and in the degree of uncertainty σ^2.

Another important case is when payoff is binary. Suppose $f = v$ over some event E and $f = 0$ elsewhere. Using Equation 13, the bias in expected utility is $(\pi_f(E) - p(E))v = \left(\frac{p(E)e^{\psi v}}{1-p(E)+p(E)e^{\psi v}} - p(E)\right)v = \frac{(e^{\psi v}-1)p(E)(1-p(E))}{1+p(E)(e^{\psi v}-1)}v$. The bias thus goes up with the strength of interests $|v|$, and goes down as indifference beliefs approach certainty ($p(E) \to 0$ or $p(E) \to 1$).

There is evidence for both comparative statics. Weinstein (1980) and Sjöberg (2000) elicit beliefs over events which vary in how desirable or undesirable they are, and find more biased beliefs over events that are either strongly desirable or strongly undesirable. Mijović-Prelec and Prelec (2010) similarly find a larger bias in an experimental treatment in which interests are stronger. Mayraz (2011) elicits predictions of future prices in different price charts, and finds more bias in charts in which subjective uncertainty is high.

The importance of accurate beliefs for decisions is *not* part of the comparative statics. When stakes are, high decision makers may put more effort into collecting information, and if this information reduces subjective uncertainty, it would also reduce the bias. However, controlling for information, the magnitude of the bias is independent of its costs.[26] Wishful thinking may thus be an important factor in high-stakes decisions, despite the resulting welfare loss.

5 Belief change

The model defines a person's beliefs in relation to her *indifference beliefs*, or the beliefs she would have held if she were completely indifferent about the state of the world (Equations 12–14). Beliefs can change for one of the following two reasons: (i) the indifference beliefs change, or (ii) the magnitude of the bias changes. Assuming the coefficient of relative optimism is fixed, the magnitude of the bias changes if and only if the decision maker's interests change. Indifference beliefs change if there is a change in relevant information.

5.1 Interests

Changes to the decision maker's interests can alter beliefs even in the absence of any new relevant information. The most important reason for a change in interests is the making of a new commitment. For example, when an investor buys some financial asset, she gains an interest in the price of the asset going up. According to the model, this should cause her beliefs about the asset to

[26] The findings in Mayraz (2011) and Hoffman (2011b) are consistent with this prediction.

become more optimistic. This prediction fits many cognitive dissonance findings, such as the Knox and Inkster (1968) finding that bettors become more confident that a horse would win the race after placing a bet on the horse.[27] Section 6.1 shows that this kind of belief change can cause commitments to escalate.

Interests can also change for exogenous reasons. Consider the beliefs of optimistic parents whose child is to be allocated randomly to one of two schools: A or B. The parents want their child to be allocated to the best school, but they initially do not know what school their child would attend. Consequently, when they learn that their child is to be allocated to school A, they gain an interest in A being the better school. The prediction of the model is that this change in interests would cause their beliefs to shift, so that they come to think more highly of school A.[28]

5.2 Relevant information

The observation of relevant new information results in a Bayesian update to the decision maker's indifference beliefs. Because the bias can itself be seen as a Bayesian update (Section 4.1), the relationship between *ex-post* distorted beliefs and *ex-ante* distorted beliefs is also Bayesian.[29]

[27] Suppose the bettor is optimistic ($\psi > 0$), and let E denote the event that the horse wins the race. Assuming no existing interest in the horse winning the race, the ex-ante payoff function is constant, and beliefs coincide with the indifference probability measure p. Placing the bet causes the payoff-difference between E and $\bar{E}$ to increase to some positive amount b. According to Equation 13, the odds-ratio between the two events increases by a factor of $e^{\psi b} > 1$.

[28] If the parents are pessimistic, their beliefs would shift in the opposite direction.

[29] Let e denote the new evidence. Inserting Bayes Rule in Equation 14 we obtain that

$$\begin{aligned}
\log \frac{\pi_{f_{\text{post}}}(A)}{\pi_{f_{\text{post}}}(B)} &= \log \frac{p_{\text{post}}(A)}{p_{\text{post}}(B)} + \psi[f(A) - f(B)] \\
&= \left(\log \frac{p_{\text{pre}}(A)}{p_{\text{pre}}(B)} + \log \frac{L(e|A)}{L(e|B)} \right) + \psi[f(A) - f(B)] \\
&= \left(\log \frac{p_{\text{pre}}(A)}{p_{\text{pre}}(B)} + \psi[f(A) - f(B)] \right) + \log \frac{L(e|A)}{L(e|B)} \\
&= \log \frac{\pi_{f_{\text{pre}}}(A)}{\pi_{f_{\text{pre}}}(B)} + \log \frac{L(e|A)}{L(e|B)}.
\end{aligned}$$

f	I	U
S	+1	0
F	−1	0

p	I	U
S	1/4	1/4
F	1/4	1/4

π_f	I	U
S	4/9	2/9
F	1/9	2/9

Figure 2: Merger example. Let S, F, I, and U denote respectively the event that the deal is successful, unsuccessful, important and unimportant. The payoff f is 1 if the deal is successful and important for promotion, -1 if it is important and unsuccessful, and 0 if it is not important. The indifference beliefs p are symmetric, and the distorted beliefs π_f are computed on the assumption that $\psi = \log 2$. Learning that the deal is important, increases the subjective probability that it is a success from 2/3 to 4/5, even though the two events are objectively independent. Similarly, learning that the deal has failed, decreases the subjective probability that it would be important from 5/9 to 1/3.

However, because distorted beliefs depend on the payoff-function, when two variables are complements or substitutes in the payoff-function, they become dependent in the decision maker's beliefs. Consequently, the Bayesian update following news about one of the two will alter beliefs about the other, even if they are objectively independent. The update in the decision maker's beliefs after observing new information is thus formally Bayesian, but may nonetheless appear biased to outside observers.

Consider the following example. An optimistic manager's promotion may or may not be dependent on the success of a merger deal, this being determined independently of the deal's success. The manager's payoff is particularly high (low) if the deal is both successful and important for promotion (unsuccessful and important). Consequently, the subjective probability that the merger is both important and successful is biased upward, and the probability that it is important and unsuccessful is biased downward. The two events are thus subjectively correlated, and news about one will affect beliefs about the other (Figure 2).

In this example, two complements (importance and success) become positively correlated in the decision maker's beliefs. Substitutes would become negatively correlated. For example, if a company pursues two research approaches in parallel, success in one would decrease the subjective probability that the other approach could have worked, whereas failure would increase the confidence that the other approach would succeed. These two effects are reversed if a decision maker is a pessimist.

The following proposition is a formal statement of these observations. Payoff is assumed to be the function of two random variables X and Y, which an unbiased decision maker would consider to be independent. An event E is observed, where E is independent of X, and is indicative of a high value of Y. Normatively, therefore, the observation of E should change beliefs about Y, but leave beliefs about X unchanged. The possibility that X and Y are complements (substitutes) is captured by the notion of supermodularity (submodularity). When that is the case, the observation of E would nonetheless result in a change in beliefs about X.

Proposition 4. *Suppose the payoff-function f is a function of two real-valued random variables X and Y, such that $p(X = x, Y = y) = p(X = x)p(Y = y)$ for all $x, y \in \mathbb{R}$, and suppose that E is an event such that $p(X = x, Y = y|E) = p(Y = y|E)$ for all x and y, and that $p(E|Y = y)$ is an increasing function of y, then*

1. $\Pi_{X|E} \succeq_{LR} \Pi_X$ *if (i)* $\psi \geq 0$ *and f is supermodular, or (ii)* $\psi \leq 0$ *and f is submodular.*

2. $\Pi_X \succeq_{LR} \Pi_{X|E}$ *if (i)* $\psi \geq 0$ *and f is submodular, or (ii)* $\psi \leq 0$ *and f is supermodular.*

Moreover, the above relations of stochastic dominance in the likelihood ratio are strict whenever $\psi \neq 0$, f is strictly supermodular/submodular, $p(E|Y = y)$ is strictly increasing in y, and neither X nor Y is almost everywhere constant.

6 Applications

In this section I explore two applications. The first is to investing, and the second to crime deterrence.

6.1 Increasing exposure to risk

Since beliefs depend on interests, a decision to bet on some event causes a change in the subjective probability for the event in question (Section 5.1). It follows that an optimistic investor who invests in a risky asset would subsequently prefer to increase her investment. Assuming the original choice is optimal, the opportunity to revise the investment leads to welfare loss. Moreover, it is possible for the investor to end up with a lower level of expected utility than if she had kept all her money in the safe asset. The investor may thus be better off without any access to the risky asset.

Consider an optimistic investor with log utility and initial wealth w who can invest a fraction α of her wealth in a risky asset that pays 1 in the good state G and -1 in the bad state B. She makes an initial investment in period 1, and can then revise her investment in period 2. If the subjective probability of the good state is q, she would choose to invest a fraction $\alpha(q) = 2q - 1$ of her wealth in the risky asset.[30]

Let $p(G) > 0.5$ denote the objective probability of the good state, and suppose that the investor has rational expectations if she has no stake in the risky asset.[31] Since this is her situation prior to the $t = 1$ decision, she would invest a fraction $\alpha_1 = 2p(G) - 1$ of her wealth in the risky asset. Following this investment, her payoff-function is $f(G) = \log(w + \alpha w) = \log(2p(G)w)$ and $f(B) = \log(w - \alpha w) = \log(2p(B)w)$. The new subjective beliefs can be computed using Equation 13:

$$\begin{aligned}\pi_f(G) &= \frac{p(G)e^{\psi f(G)}}{p(G)e^{\psi f(G)} + p(B)e^{\psi f(B)}} = \frac{p(G)e^{\psi \log(2p(G)w)}}{p(G)e^{\psi \log(2p(G)w)} + p(B)e^{\psi \log(2p(G)w)}} \\ &= \frac{p(G)(2p(G)w)^{\psi}}{p(G)(2p(G)w)^{\psi} + p(B)(2p(B)w)^{\psi}} = \frac{1}{1 + \left(\frac{p(B)}{p(G)}\right)^{1+\psi}} > p(G),\end{aligned} \tag{18}$$

where the inequality follows from the assumption that $\psi > 0$ (the investor is optimistic). Given that $\pi_f(G) > p(G)$, the investor would increase the share invested in the risky asset to $\alpha_2 = 2\pi_f(G) - 1 > \alpha_1$. Since expected utility is a strictly concave function of this amount, the opportunity to revise the investment results in welfare loss. Moreover, $\lim_{\psi\to\infty} \pi_f(G) = 1$, and hence $\lim_{\psi\to\infty} \alpha = 1$. Furthermore, $\lim_{x\to 0} \log(x) = -\infty$. Combining these observations, it follows that as long as $p(G) < 1$, there is a critical value ψ^*, such that for all $\psi > \psi^*$ the (objective) expected utility following the opportunity to revise the investment is below the utility of investing nothing in the risky

[30]Expected utility is $q\log(w + \alpha w) + (1 - q)\log(w - \alpha w)$. Since $q > 0.5$, the solution is internal. Solving the first order condition we obtain that $\alpha(q) = 2q - 1$.

[31]The rational expectations assumptions is useful for welfare analysis.

asset. A sufficiently optimistic investor would thus be better off without the opportunity to invest in the risky asset.

6.2 Deterrence

The two principal approaches to crime deterrence are (i) improving law enforcement, and (ii) increasing sentencing. The first makes punishment more *likely*, and the second makes it more *severe*. There is good evidence that the first approach is considerably more effective Grogger, 1991; Nagin and Pogarsky, 2001; Durlauf and Nagin, 2011. In this section I offer one explanation for why that may be the case. I follow Becker (1968) in modeling the decision to engage in crime as rational, but assume that criminals are optimistic, and that they therefore underestimate the probability that they would end up in jail. An increase in the severity of punishment increases the payoff difference between getting caught and not getting caught. As I show in the formal model, this leads to an an increase in the bias. Thus, a more severe punishment has an ambiguous effect on deterrence: on the one hand jail is subjectively *worse* (the sentence is more severe), but on the other hand it is subjectively *less likely* (because of the increased optimistic bias). In some cases, an increase in the severity of punishment can even be counter-productive. By contrast, increasing the likelihood that crime is punished leaves the bias in beliefs unchanged, and unambiguously improves deterrence.

An optimistic criminal has to choose whether to continue a life of crime or to take up a job at McDonald's. There are two states corresponding to whether or not crime would land the criminal in jail. The payoff from crime is $f(B) = -c$ in the bad state and $f(G) = 0$ in the good state. The payoff from a job at McDonald's is $-b$ in both states, with $0 < b < c$. Let p denote the probability measure representing the beliefs the criminal would have had if she were indifferent as to whether crime would land her in jail, and assume that p is unbiased.[32] Since crime is the status-quo, subjective beliefs are represented by the distorted probability measure π_f, where f denotes the payoff-function of a criminal. Using Equation 13 we obtain that $\frac{\pi_f(B)}{\pi_f(G)} = \frac{p(B)}{p(G)} e^{-\psi c}$. Deterrence is successful if the expected gain from quitting crime is more than the expected

[32]The minimal assumption is that improving law enforcement increases $p(B)$.

loss. This is the case if $\pi_f(B)(c - b) \geq \pi_f(G)b$, or

$$\frac{b}{c-b} \leq \frac{\pi_f(B)}{\pi_f(G)} = \frac{p(B)}{p(G)}\, e^{-\psi c}. \tag{19}$$

Consider the following two potential policy changes. First, the government can improve law enforcement, thereby increasing $\frac{p(B)}{p(G)}$. Holding c constant, such a change would increase the RHS of Equation 19, while leaving the LHS unchanged, and would therefore improve deterrence for any level of optimism. Second, the government can increase in the severity of jail c, leaving its probability unchanged. If $\psi = 0$ the change would reduce the LHS, and leave the RHS unchanged. Thus, for realist criminals any increase in the severity of punishment improves deterrence. However, for optimistic criminals $\psi > 0$, and so the increase in c reduces the RHS of Equation 19 at the same time as it is reducing the LHS. There are thus two forces pulling in opposite directions: (i) the *utility effect* works to *increase* deterrence (jail is worse), and (ii) the *probability effect* works to *decrease* deterrence (jail is subjectively less likely). Since $\lim_{c\to\infty}(c - b)e^{-\psi c} = 0$ for $\psi > 0$, making the punishment more severe is counter-productive beyond a certain point (Figure 3). Note that the key to these results is the assumption that crime is the status-quo. A decision maker who has never before engaged in crime *would* be deterred by a more severe punishment.

7 Conclusion

By tying beliefs to the decision maker's existing interests—rather than to the choices that she faces—the model separates optimism and pessimism from attitudes toward risk and ambiguity. Since beliefs depend on interests, any change to these interests leads to a change in beliefs. This allows the model to capture not only static belief biases, but also the dynamic phenomenon of cognitive dissonance. Since the bias is not assumed to be welfare enhancing, it is possible to model pessimistic as well as optimistic bias, and to model optimistic bias in situations where it leads to costly mistakes. The model is tractable and parsimonious, and can be used both in the construction of entirely new applications, and in adapting existing applications to incorporate the implications of wishful thinking, cognitive dissonance, and other related biases.

While the model is surely too simple to be empirically correct, it does seem consistent with broad features of the evidence. In particular, the existing evi-

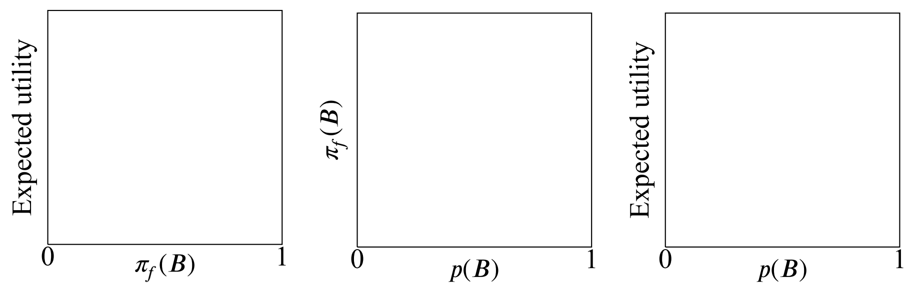

Figure 3: The impact on the subjective utility of crime of increasing punishment levels from relatively lenient (solid blue line) to severe (dotted red line). At any given subjective probability, greater severity reduces utility (panel 1). However, greater severity also reduces the subjective probability that the bad state is realized (panel 2). The net effect (panel 3) is ambiguous, and can actually be positive if the objective probability of the bad state is low.

dence suggests that the magnitude of the bias is indeed not directly dependent on its cost Mayraz, 2011; Hoffman, 2011b. This has the important implication that the biases captured by the model are not limited to low stakes decisions, and may well be important in high-stakes decision environments. When stakes are high, decision makers have strong incentives to double-check their probability judgments, and in particular to try and detect any evidence of bias in their beliefs. However, whenever there is significant uncertainty, there is a large range of plausible views, and biased beliefs would generally fall within this range. For this reason, even highly motivated and sophisticated decision makers may be unable to determine whether their own beliefs are biased, and would not be able to prevent such bias from affecting their decisions.

Both individual decision makers and policy makers may, however, be able to identify *situations* in which biased beliefs are liable to be a significant problem. Decision makers may then try to either reduce the degree of bias, or avoid such situations altogether. The deterrence application is an example of the former strategy: policy makers, who realize that criminals may be biased in judging the probability of ending up in jail, may therefore prefer to impose less severe punishments, and put their resources instead into increasing the probability that criminals are brought to justice. In the investment application, sophisticated investors may adopt the second strategy, committing to some particular portfolio, and not allowing themselves the flexibility of revising it. Such a strategy involves a trade-off analogous to the bias-variance trade-off in statistics: flexibility makes it possible to use more information in decision making, but at the same time it opens up the door for bias.

One important weakness of the model is the need to specify exactly what elements of uncertainty are subject to belief distortion. For example, the predictions in the deterrence application (Section 6.2) depend crucially on the assumption that the severity of punishment is taken as given, whereas the likelihood of getting caught is subjective.[33] A second weakness is that in some situations the decision maker's interests are tied to her plans, and these plans may not be readily observable. Consider two investors who are holding the same portfolio. A short-term speculator stands to gain if the portfolio does well over the near future, and would therefore be biased about this possibility. A long-term investor would, instead, be biased about long-term performance. Making a prediction therefore requires the ability to identify the type of the investor, as well as the stocks that she owns.

This paper assumes throughout that the coefficient of relative optimism is a stable characteristic of a person. One intriguing possibility, however, is that it increases following good events, and decreases following bad ones. Consistent with this idea, there is evidence that people are more positive about the stock market when the weather is good Saunders, 1993; Hirshleifer and Shumway, 2003, or after their soccer team wins an international match Edmans et al., 2007. Similarly, optimistic bias in Google's prediction markets is particularly high in days in which Google's stock is appreciating Cowgill et al., 2009. Such dynamics may be important during the popping of an asset bubble, when decreasing prices lead to losses, which in turn (if this hypothesis is true) reduce the optimistic bias among investors. A reduction in optimistic bias would then lead to further selling and further drops in prices.

A Proofs

Lemma 1

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:

Part 1. Let a denote some arbitrary constant payoff-function. Define $p = \pi_a$,

[33] If these assumptions were reversed, an increased likelihood of punishment would lead to increased bias over the severity of punishment.

and let $\mathcal{S}^* = \text{f}A \in S : p(A) > 0\text{g}$. Define $h_f(A) = \frac{\pi_f(A)}{p(A)}$ for $A \in \mathcal{S}^*$ and $h_f(A) = 0$ for $A \notin \mathcal{S}^*$. For $A \in \mathcal{S}^*$ the claim follows from the definition of h_f. By Absolute Continuity $p(A) = 0 \Rightarrow \pi_f(A) = 0$, and hence the claim holds also for $A \notin \mathcal{S}^*$. $\square$

Part 2. Let $A \in \mathcal{S}^*$ and $x \in X$, let $f(A, x)$ be the payoff-function mapping A to x and all states outside A to a. Let $E_1, \dots, E_n$ denote the other events in $\mathcal{S}^*$. By Minimal Complexity and Absolute Continuity $\mathcal{S}^*$ includes at least two events other than A. $f(A, x)$ and the constant payoff-function a agree on E_i and E_j for all i and j. Hence, by Consequentialism with $E = E_i \cup E_j$, $\frac{\pi_{f(A,x)}(E_i)}{\pi_{f(A,x)}(E_j)} = \frac{p(E_i)}{p(E_j)}$. Thus,

$$1 - \pi_{f(A,x)}(A) = \sum_i \pi_{f(A,x)}(E_i) = \sum_i \frac{\pi_{f(A,x)}(E_j)}{p(E_j)} p(E_i) = \frac{\pi_{f(A,x)}(E_j)}{p(E_j)}(1 - p(A)). \tag{20}$$

Define $\mu_A(x) = \left(\frac{1-p(A)}{p(A)}\right)\left(\frac{\pi_{f(A,x)}(A)}{1-\pi_{f(A,x)}(A)}\right)$. By Equation 20,

$$p(A)\mu_A(f(A)) = (1 - p(A))\frac{\pi_{f(A,f(A))}(A)}{1 - \pi_{f(A,f(A))}(A)} = p(E_j)\,\frac{\pi_{f(A,f(A))}(A)}{\pi_{f(A,f(A))}(E_j)}. \tag{21}$$

Let f be any payoff-function, and let A and B be any two events in $\mathcal{S}^*$. Let f' be a payoff-function that coincides with f on A and B, and with a elsewhere, and let C be any third event in $\mathcal{S}^*$. Inserting $E_j = C$ in Equation 21 we obtain that

$$\begin{aligned} \frac{\pi_f(A)}{\pi_f(B)} = \frac{\pi_{f'}(A)}{\pi_{f'}(B)} = \frac{\pi_{f'}(A)/\pi_{f'}(C)}{\pi_{f'}(B)/\pi_{f'}(C)} &= \frac{p(C)}{p(C)}\,\frac{\pi_{f(A,f(A))}(A)/\pi_{f(A,f(A))}(C)}{\pi_{f(B,f(B))}(B)/\pi_{f(B,f(B))}(C)} \\ &= \frac{p(A)\mu_A(f(A))}{p(B)\mu_B(f(B))} \end{aligned} \tag{22}$$

where the first and third steps follows from Consequentialism, and the final step from Equation 21. Since Equation 22 holds for all $A, B \in \mathcal{S}^*$ it follows that Equation 5 holds for any event $A \in \mathcal{S}^*$. For an event $A \notin \mathcal{S}^*$, define $\mu_A(x) = 1$ for all x. Since $\pi_f(A) = p(A) = 0$ for $A \notin \mathcal{S}^*$ Equation 5 holds however μ_A is defined. Combining these results Equation 5 holds for any payoff-function f and any event $A \in \mathcal{S}$. $\square$

Part 3. Let $A^* \in \mathcal{S}^*$ be some event. Define the mapping $v : X \to \mathbb{R}_+$ by $v(x) = \mu_{A^*}(x)$. For $x \in X$ let x denote also the constant payoff-function yielding the payoff x in all states. Inserting $f = x$ and $B = A^*$ in Equation 22 we obtain that for all $A \in \mathcal{S}^*$ and $x \in X$, $\frac{\pi_x(A)}{\pi_x(A^*)} = \frac{p(A)}{p(A^*)}\frac{\mu_A(x)}{v(x)}$. Since x is a constant payoff-function it follows from Indifference that $\pi_x = \pi_a = p$. Hence, $\mu_A(x) = v(x)$. Thus, $\pi_f(A) \propto p(A)v(f(A))$ for all $A \in \mathcal{S}^*$. Finally, this is also trivially true for $A \notin \mathcal{S}^*$, since $\pi_f(A) = p(A) = 0$ for $A \notin \mathcal{S}^*$. □

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 payoff-functions f_x and $g_{x,y}$ as follows: $f_x(s) = x$ for $s \in A$ and $f_x(s) = 0$ for $s \notin A$, and $g_{x,y} = f_x + y$. By Shift-Invariance, $\pi_{g_{x,y}} = \pi_{f_x}$, and in particular $\frac{\pi_{g_{x,y}}(A)}{\pi_{g_{x,y}}(B)} = \frac{\pi_{f_x}(A)}{\pi_{f_x}(B)}$. By Equation 6 it follows that $\frac{v(x+y)}{v(y)} = \frac{v(x)}{v(0)}$. Hence, defining $\sigma(x) = \log(\frac{v(x)}{v(0)})$ we obtain that σ is linear, i.e. for all x and y, $\sigma(x + y) = \sigma(x) + \sigma(y)$. For $m \in \mathbb{N}$ let $y = mx$. By induction we obtain that $\sigma(mx) = m\sigma(x)$. Similarly, for $n \in \mathbb{N}$ let $y = \frac{x}{n}$ to obtain that $\sigma(x) = \sigma(ny) = n\sigma(y)$, and hence $\sigma(\frac{x}{n}) = \frac{\sigma(x)}{n}$. Let $y = -x$ to obtain that $\sigma(-x) = -\sigma(x)$. Combining these results, and defining $\psi = \sigma(1)$, we obtain that for any rational number $q \in X$, $\sigma(q) = \psi q$, and so $v(q) = v(0)e^{\psi q}$. Let now $x \in X$ be any feasible payoff-value, and let fq_ng$_{n\in\mathbb{N}}$ be a sequence of rational feasible payoff-values converging to x. By prize-continuity $\pi_{f_{q_n}} \to \pi_{f_x}$, which given Equation 6 implies that $v(q_n) \to v(x)$. By the result for rational numbers, $v(q_n) = v(0)e^{\psi q_n}$, and hence $v(q_n) \to v(0)e^{\psi x}$. Thus, $v(x)$ and $v(0)e^{\psi x}$ are both the limit of the same sequence of real-numbers, and so $v(x) = v(0)e^{\psi x}$. Finally, since Equation 6 is invariant to multiplying v by a positive number, we obtain that Equation 7 holds for all $x \in \mathbb{R}$. □

Lemma 2

Proof. Let $a \in F$ denote some constant payoff-function, and define $p = \pi_a$. 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 $\pi_f(A) > 0$ for any $f \in F$ and $A \in \mathcal{S}$. Let $\Sigma(\mathcal{S}) \subseteq \Sigma$ denote the algebra generated by $\mathcal{S}$, and let $F(\mathcal{S}) \subseteq F$ denote the set of $\Sigma(\mathcal{S})$-measurable payoff-functions. By Lemma 1 there exists a probability measure $p_{\mathcal{S}}$ over $(S, \Sigma(\mathcal{S}))$ and a parameter $\psi_{\mathcal{S}} \in \mathbb{R}$ such that Equation 7 holds any probability measure $f \in F(\mathcal{S})$ and any event $A \in \mathcal{S}$. In particular $a \in F(\mathcal{S})$ (any constant payoff-function is), and hence for any $A \in \mathcal{S}$, $p(A) = \pi_a(A) \propto p_{\mathcal{S}}(A)e^{\psi_{\mathcal{S}} a}$. Thus, $p(A) = p_{\mathcal{S}}$ for any event

$A \in \mathcal{S}$, and hence also for any event $A \in \Sigma(\mathcal{S})$. Define $\psi = \psi_{\mathcal{S}}$. It follows that for any payoff-function $f \in \Sigma(\mathcal{S})$ and any event $A \in \mathcal{S}$, $\pi_f(A) \propto p(A)e^{\psi f(A)}$.[34]

Let now A and B denote any events such that $p(B) > 0$, and let f be any payoff-function. I need to show that $\frac{\pi_f(A)}{\pi_f(B)} = \frac{p(A)}{p(B)} e^{\psi(f(A)-f(B))}$. To simplify notation let $\delta_f(A, B) = \log \frac{\pi_f(A)}{\pi_f(B)} - \log \frac{p(A)}{p(B)}$. With this notation I need to prove that $\delta_f(A, B) = \psi(f(A) - f(B))$. Let $E_1, E_2, \dots E_n$ denote the events in $\mathcal{S}$. Without limiting generality suppose $A \cap E_1$ is not-null. Define a payoff-function $g \in F$ by $g = f(A)$ on $A \cap E_1$ and $g = f(B)$ elsewhere, and a payoff-function $h \in F(\mathcal{S})$ by $h = f(A)$ on E_1 and $h = f(B)$ elsewhere. With these definitions, $\delta_f(A, B) = \delta_f(A \cap E_1, B) = \delta_g(A \cap E_1, B) = \delta_g(A \cap E_1, B \cup E_2) = \delta_g(A \cap E_1, E_2) = \delta_h(A \cap E_1, E_2) = \delta_h(E_1, E_2) = \psi(f(A) - f(B))$, where the last step uses the fact that h is in $F(\mathcal{S})$, and the other steps use Consequentialism and the fact that by Shift-Invariance $\pi_{f(A)} = \pi_{f(B)} = p$. □

Corollary 1

Proof. The proof that a logit distortion is well-behaved is trivial. I thus prove only that if π is well-behaved then it is a logit-distortion. The conditions of Lemma 2 are met. Let p and ψ be parameters for which the claim in Lemma 2 holds. Suppose f is a.e. simple then there exist a finite set of disjoint events $\{E_1, \dots, E_n\}$ such that f is constant on each of these events, and for some payoff-function g, $\pi_g(\cup_i E_i) = 1$. By Absolute Continuity also $\pi_f(\cup_i E_i) = 1$, and so $\pi_f(A \cap \cup_i E_i) = \pi_f(A)$. Given that the events are disjoint it follows that $\pi_f(A) = \sum_i \pi_f(A \cap E_i)$. Using Lemma 2 we obtain that $\pi_f(A) \propto \sum_i p(A \cap E_i)e^{\psi f(A \cap E_i)}$. By Absolute Continuity $p(S \setminus \cup_i E_i) = 0$, and hence $\int_A e^{\psi f} \, \mathrm{d}p = \sum_i p(A \cap E_i)e^{\psi f(A \cap E_i)}$. Combining these observations we obtain that $\pi_f(A) \propto \int_A e^{\psi f} \, \mathrm{d}p$. □

Lemma 3

Proof. The case of $\psi = 0$ is trivial. Henceforth I assume $\psi \neq 0$. By Corollary 1 there exist a probability measure p and a parameter ψ such that Equation 3 holds for any payoff-function f that is almost everywhere simple. If there exists a payoff-function f that is not almost everywhere simple then

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

there exists an infinite sequence $\{A_n\}_{n\in\mathbb{N}}$ of disjoint non-null events.[35] I need to prove that in this case there exists a number $M \in \mathbb{R}$ such that $e^{\psi x} \leq M$ 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, $p(A_n)e^{\psi x_n} \geq p(A_1)e^{\psi x_1}$. Define a payoff-function f by $f(A_n) = x_n$ for $n \in \mathbb{N}$, and $f(s) = x_1$ outside $\cup_n A_n$. For $n \in \mathbb{N}$ define also a simple payoff-function f_n by $f_n(A_n) = x_n$ and $f_n(s) = x_1$ for $s \notin A_n$. By construction f and f_n agree on A_1 and A_n. Thus, for all $N \in \mathbb{N}$, $1 \geq \sum_{n\leq N} \pi_f(A_n) = \pi_f(A_1)\sum_{n\leq N} \frac{\pi_f(A_n)}{\pi_f(A_1)} = \pi_f(A_1)\sum_{n\leq N} \frac{\pi_{f_n}(A_n)}{\pi_{f_n}(A_1)} = \pi_f(A_1)\sum_{n\leq N} \frac{p(A_n)e^{\psi x_n}}{p(A_1)e^{\psi x_1}} \geq \pi_f(A_1)\sum_{n\leq N} 1 = N\pi_f(A_1)$ where the second equality follows from Consequentialism, and the third from Corollary 1. Letting $N \to \infty$ we obtain that $\pi_f(A_1) = 0$, in contradiction to the assumption that A_1 is not null. □

Lemma 4

Proof. By Corollary 1 there exist a probability measure p and a parameter ψ such that Equation 9 holds for any payoff-function f that is almost everywhere simple. I show that the claim holds with the same p and ψ also for a payoff-function f that is not everywhere simple. If such payoff-functions then by Lemma 3 there exists a number M, such that $e^{\psi x} \leq M$ for all $x \in X$. Assume first that $\psi \neq 0$. For any $n \in N$ divide the interval $[m, M]$ into 2^n non-overlapping intervals of length $\frac{M-m}{2^n}$. For any state s let $I_n(s)$ denote the interval to which $e^{\psi f}$ belongs, and let $I_n^{min}(s)$ denote its lower endpoint. Define a simple payoff-function f_n by $f_n(s) = \log \frac{I_n^{min}(s)}{\psi}$. With this definition $e^{\psi f(s)} - \frac{M-m}{2^n} \leq e^{\psi f_n(s)} \leq e^{\psi f(s)}$ for all s, and so $e^{\psi f_n(s)} \nearrow e^{\psi f(s)}$ for all s. Moreover, since $e^{\psi x}$ is a monotonic continuous function also $f_n \to f$. Thus,

$$\frac{\pi_f(A)}{\pi_f(B)} = \frac{\lim \pi_{f_n}(A)}{\lim \pi_{f_n}(B)} = \lim \frac{\pi_{f_n}(A)}{\pi_{f_n}(B)} = \lim \frac{\int_A e^{\psi f_n}\,\mathrm{d}p}{\int_B e^{\psi f_n}\,\mathrm{d}p} = \frac{\lim \int_A e^{\psi f_n}\,\mathrm{d}p}{\lim \int_B e^{\psi f_n}\,\mathrm{d}p} = \frac{\int_A e^{\psi f}\,\mathrm{d}p}{\int_B e^{\psi f}\,\mathrm{d}p} \tag{23}$$

[35] If f 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 f is almost always simple. Since f has no atoms on E it follows that there exists a value y (the median of f on E) such that $p(s \in E : f(s) \leq y) = \frac{p(E)}{2}$. Thus E includes two non-null events on which f 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.

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

Theorem 1

Proof. I prove that if π is a well-behaved distortion then it is a logit distortion. The opposite direction is trivial. By Lemma 4 there exist a probability measure p and a parameter ψ such that Equation 9 holds for any events A and B for which $p(B) > 0$ and any payoff-function f for which there exists a number $m > 0$ such that $e^{\psi f} \geq m$ on $A \cup B$. To complete the proof I need to show that Equation 9 holds even if no such number m exists. Let f be any payoff-function and let A and B be any events such that $p(B) > 0$. If f is almost everywhere simple the claim follows from Corollary 1. Otherwise, by Lemma 3 there exists a number $M \in \mathbb{R}$ such that $e^{\psi x} \leq M$ for all $x \in X$. For $n \in \mathbb{N}$ let $A_n = \{s \in A : e^{\psi f} \geq 2^{-n}\}$, and similarly define B_n. By construction $\lim_{n\to\infty} A \setminus A_n = \varnothing$ and similarly $\lim_{n\to\infty} B \setminus B_n = \varnothing$. Moreover, since $p(B) > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $p(B_n) > 0$. 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

$$\frac{\pi_f(A)}{\pi_f(B)} = \lim_{n\to\infty} \frac{\pi_f(A_n)}{\pi_f(B_n)} = \lim_{n\to\infty} \frac{\int_{A_n} e^{\psi f}\, \mathrm{d}p}{\int_{B_n} e^{\psi f}\, \mathrm{d}p} = \frac{\lim_{n\to\infty} \int_{A_n} e^{\psi f}\, \mathrm{d}p}{\lim_{n\to\infty} \int_{B_n} e^{\psi f}\, \mathrm{d}p} = \frac{\int_A e^{\psi f}\, \mathrm{d}p}{\int_B e^{\psi f}\, \mathrm{d}p} \tag{24}$$

where step 3 holds since the integrals are bounded from below and above: (i) $e^{\psi x} \leq M$ for all $x \in X$, so the integrals are bounded from above by M, and (ii) $p(B_{n_0}) > 0$ and $f \geq 2^{-n_0}$ on B_{n_0}, and hence there exists some $m > 0$ such that for all $n \geq n_0$, $\int_{B_n} e^{\psi f}\, \mathrm{d}p \geq \int_{B_{n_0}} e^{\psi f}\, \mathrm{d}p \geq 2^{-n_0} p(B_{n_0}) \geq m$. $\square$

Lemma 6

Proof. Starting with A1, let f and f' be payoff-functions and E an event, and suppose that $\pi_f(E) = 0$. I need to prove that $\pi_{f'}(E) = 0$. Let e and e' be acts such that $f = ue$ and $f' = ue'$. Since $\pi_f(E) = 0$ it follows from Equation 11 that for any acts g and h that differ only in E, $V_e(g) = V_e(h)$, and so E is e-null. By B6 it follows that E is also e'-null. Let a and b be constant acts such that $a \succ_{e'} b$, and let g be an act defined by $g = b$ on E and $g = a$ outside

E. By construction a and g differ only on E, and hence (since E is e'-null) $V_{e'}(a) = V_{e'}(g)$. Using Equation 11 it follows that $\pi_{f'}(E)(u(a) - u(b)) = 0$. By assumption $a \succ_{e'} b$, and hence $u(a) - u(b) > 0$. Thus, $\pi_{f'}(E) = 0$ as required.

For A2 suppose $f = f'$ over a non-null event E. I need to prove that for any event $A \subseteq E$, $\pi_f(A|E) = \pi_{f'}(A|E)$. If $\pi_f(A) = 0$ the claim follows from A1. Suppose therefore that $\pi_f(A) > 0$. Let e and $\hat{e}$ be acts such that $f = ue$ and $f' = u\hat{e}$. Define an act e' by $e' = e$ on E and $e' = \hat{e}$ outside E. By construction $ue' = ue = f$ on E and $ue' = u\hat{e} = f'$ outside E. By assumption $f = f'$ on E and hence $ue' = f$ everywhere. By B5 and B6 there exist constant acts $\overline{a}$ and $\underline{a}$ such that $\overline{a} \succeq_e g \succeq_e \underline{a}$ for any g. Define an act g^* by $g^* = \overline{a}$ on $A \cup E^c$ and $g^* = \underline{a}$ on $E \setminus A$. For any constant act b, define an act g_b by $g_b = b$ on E and $g_b = \overline{a}$ outside E. By Equation 11, $g^* \succeq_e g_b$ if and only if $\pi_f(A)u(\overline{a}) + \pi_f(E \setminus A)u(\underline{a}) \geq \pi_f(E)u(b)$, or equivalently, $\pi_f(A|E)u(\overline{a}) + (1 - \pi_f(A|E))u(\underline{a}) \geq u(b)$. Since the set of constant acts is closed under arbitrary mixing, there exists a constant act b^* such that $\pi_f(A|E)u(\overline{a}) + (1 - \pi_f(A|E))u(\underline{a}) = u(b^*)$, and equivalently, $g^* \sim_e g(b^*)$. Since $e = e'$ on E, and since g^* and $g(b^*)$ differ only on E, it follows from B7 that also $g^* \sim_{e'} g(b^*)$, and therefore also $\pi_{f'}(A|E)u(\overline{a}) + (1 - \pi_{f'}(A|E))u(\underline{a}) = u(b^*)$. Combining these results we obtain that $\pi_f(A|E)(u(\overline{a}) - u(\underline{a})) = \pi_{f'}(A|E)(u(\overline{a}) - u(\underline{a}))$. Finally, since $a \succ_e \underline{a}$ then $u(\overline{a}) > u(\underline{a})$, and so $\pi_{f'}(A|E) = \pi_f(A|E)$ as required.

For A3 suppose $f' = f + x$ for some real number x. I need to show that $\pi_{f'} = \pi_f$. Let e and e' be acts, such that $f = ue$ and $f' = ue'$. By B4 and B5 there exists constant acts a and b, such that $a \succ_e b$ and $a \succeq_e g \succeq_e b$ for any act g. Let $\Delta u = u(a) - u(b)$. Define a function $\beta : S \to \mathbb{R}$ by $\beta = \frac{f - u(b)}{\Delta u - x} \frac{x}{\Delta u}$. Since by construction $a \succeq_e g \succeq_e b$ for any act g it follows that for any s, $u(a) \geq f(s) + x \geq u(b)$, and hence $f(s) - u(b) \leq \Delta u - x$ and $x \leq \Delta u$. Thus, $0 \leq \beta(s) \leq 1$ for all s. Define an act g by $g = (1 - \beta)e + \beta e'$, let $\alpha = \frac{x}{\Delta u}$, and let h and h' be acts defined by $h = (1 - \alpha)g + \alpha b$ and $h' = (1 - \alpha)g + \alpha a$. With these definitions,

$$
\begin{aligned}
uh &= (1 - \alpha)ug + \alpha u(b) = (1 - \alpha)u((1 - \beta)e + \beta e') + \alpha u(b) \\
&= (1 - \alpha)((1 - \beta)f + \beta(f + x)) + \alpha u(b) = (1 - \alpha)f + \alpha u(b) + (1 - \alpha)\beta x \\
&= (1 - \alpha)f + \alpha u(b) + \left(\frac{\Delta U - x}{x}\right)\left(\frac{f - u(b)}{\Delta u - x}\right)\left(\frac{x}{\Delta u}\right) x \\
&= (1 - \alpha)f + \alpha u(b) + \alpha(f - u(b)) = f,
\end{aligned}
$$

and $uh' = uh + \alpha\Delta u = f + x = f'$. By B8, $\succeq_h = \succeq_{h'}$. Hence by Lemma 5 and the result just obtained we conclude that $\pi_f = \pi_{f'}$.

For A4 suppose that $f_n \to f$ uniformly. I need to prove that for any event E and $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for $n > n_0$, $|\pi_{f_n}(E) - \pi_f(E)| < \epsilon$. By B5 there exist constant acts a and b, such that $a \succeq_e g \succeq_e b$ for any act g. In particular, for any state s, there exists $\beta : S \to [0,1]$ such that $f = (1-\beta)u(b) + \beta u(a)$. Moreover, for any n, there exists $\beta_n : S \to [0,1]$ such that $f_n = (1-\beta_n)u(b) + \beta_n u(a)$.

Let $e = (1-\beta)b + \beta a$ and for any n let $e_n = (1-\beta_n)b + \beta_n a$. By construction, $f = ue$ and for any n, $f_n = ue_n$. Let $\delta > 0$ be some number. Since $f_n \to f$ uniformly there exists $N > 0$ such that for any $n > N$ and for any state s, $|f_n(s) - f(s)| < \delta$. Thus, for any s, $|u(e(s)) - u(e_n(s))| < \delta$.

Let $\Delta u = u(a) - u(b)$. By construction, $u(e(s)) - u(e_n(s)) = (\beta(s) - \beta_n(s))\Delta u$. Hence, for any $n > N$ and any state s, $|\beta(s) - \beta_n(s)| < \frac{\delta}{\Delta u}$. In particular, for any final outcome $z \in Y$, $|e(s)(z) - e_n(s)(z)| < \frac{\delta}{\Delta u}$. Since this is true for any $\delta > 0$ it follows that $e_n \to e$ uniformly. Hence, by B9 $\succeq_{e_n} \to \succeq_e$.

Define an act g^* by $g^*(E) = a$ and $g^*(S \setminus E) = b$. By Equation 11 for any constant act c, $c \succeq_e g^*$ if and only if $\pi_f(E) \le \frac{u(c)-u(b)}{u(a)-u(b)}$. Let c^* be a constant act such that $g^* \sim_e c^*$. Thus, for any constant act c such that $u(c) > u(c^*)$, $c \succ_e g^*$, since $\succeq_{e_n} \to \succeq_e$ for any such c then for any n large enough $c \succ_{e_n} g^*$. By Equation 11 this implies that $\pi_{f_n}(E) \le \frac{u(c)-u(b)}{u(a)-u(b)}$. Since this is true for any $c > c^*$ it follows that for any $\delta > 0$ for n large enough $\pi_{f_n}(E) < \pi_f(E) + \delta$. A symmetric argument can be made for c for which $u(c) < u(c^*)$. Combining these results we obtain that for any $\delta > 0$ for n large enough $\pi_{f_n}(E) \in [\pi_f(E) - \delta, \pi_f(E) + \delta]$, and so $\pi_{f_n}(E) \to \pi_f(E)$ as required. □

Theorem 2 (uniqueness part)

Proof. Note first that the utility function is determined up to a positive affine transformation by preferences over constant Anscombe-Aumann acts. Hence, if the triplet (p', u', ψ') represents $\succeq$ there exists real numbers $\alpha > 0$ and β such that $u' = \alpha u + \beta$. Next, let e be any act and A any event. By B5 there exist constant acts a and b such that $a \succ_e b$. Let g_A be an act defined by $g_A = a$ on A and $g_A = b$ on A^c. Let c_A be a constant act defined by $c_A = \pi_{ue}(A)a + (1 - \pi_{ue}(A))b$. By Equation 10 $g_A \sim_e c_A$. Similarly, let $c'_A = \pi_{u'e}(A)a + (1 - \pi_{u'e}(A))b$ then also $g_A \sim_e c'_A$, and so $c_A \sim_e c'_A$. Since $a \succ_e b$ it follows that $u(a) > u(b)$, and hence by Equation 10 $\pi_{u'e}(A) = \pi_{ue}(A)$. Since this is true for all A it follows that for any act e, $\pi_{u'e} = \pi_{ue}$. I now use this observation to show first that $p' = p$ and then that $\psi' = \frac{\psi}{\alpha}$. First, let e be a constant act. By Equation 3 $\pi_{ue} = p$ and

$\pi_{u'e} = p'$. Hence it follows from the above observation that $p' = p$. Finally, by Minimal Complexity and B6 there exists an event A such that neither A nor A^c is $\succeq_e$ null for any act e. Applying the above observation to g_A we obtain that $\pi_{u'g_A} = \pi_{ug_A}$. By Equation 3, $\log \frac{\pi_{ug_A}(A)}{\pi_{ug_A}(A^c)} = \log \frac{p(A)}{p(A^c)} + \psi(u(a) - u(b))$. Using the corresponding equation for $\pi_{u'g_A}$, the result that $p' = p$, and the fact that $u' = \alpha u + \beta$, we obtain that $\psi(u(a)-u(b)) = \psi'(u'(a)-u'(b)) = \alpha\psi'(u(a)-u(b))$. Hence $\psi' = \frac{\psi}{\alpha}$. □

Proposition 1

Proof. By Equation 3, $\Pi_f(x) \propto \int_{-\infty}^{x} h(x)\,\mathrm{d}P_f(x)$ with $h(x) = e^{\psi x}$. If $\psi \geq 0$, $e^{\psi x}$ is non-decreasing, and hence $\Pi_f \succeq_{\mathrm{LR}} P_f$. If $\psi \leq 0$, $h' = e^{-\psi x}$ is non-decreasing, and $P_f(x) \propto \int_{-\infty}^{x} \mathrm{d}P_f(x) = \int_{-\infty}^{x} h'(x)e^{\psi x} d\,P_f(x) = \int_{-\infty}^{x} h'(x)\,\mathrm{d}\Pi_f$, and so $P_f \succeq_{\mathrm{LR}} \Pi_f$. The first part of the claim follows since stochastic dominance in the likelihood ratio implies first-order stochastic dominance. □

Proposition 2

Proof. I prove only that $\lim_{\psi\to\infty} \pi_f(A_{\max}) = 1$, as the proof that $\lim_{\psi\to-\infty} \pi_f(A_{\min}) = 1$ is very similar. Let $x_1 > x_2 > \cdots x_n$ denote the payoffs in the range of f. Thus, $A_{max} = f^{-1}(x_1)$, and by Equation 13,

$$\lim_{\psi\to\infty} \pi_f(A_{max})^{-1} = \lim_{\psi\to\infty}\left(\frac{p(f^{-1}(x_1))e^{\psi x_1}}{\sum_i p(f^{-1}(x_i))e^{\psi x_i}}\right)^{-1} = 1 + \sum_{i>1} \frac{p(f^{-1}(x_i))}{p(f^{-1}(x_1))} \lim_{\psi\to\infty} e^{\psi(x_i - x_1)},$$

and since $x_i < x_1$ for $i > 1$, the last term is zero. Hence, $\lim_{\psi\to\infty} \pi_f(A_{\max}) = 1$, as required. □

Proposition 3

Proof. By Equation 3 and the assumption that $P_X \sim \mathcal{N}(\mu, \sigma^2)$,

$$\begin{aligned}
\Pi_X(x) &\propto \int_{-\infty}^{x} e^{\psi f(x)} p(X = x)\,\mathrm{d}x = \int_{-\infty}^{x} e^{\psi(ax+b)}\left(\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)\mathrm{d}x \\
&= e^{\psi b}\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2 - 2\psi a\sigma^2 x}{2\sigma^2}}\,\mathrm{d}x = e^{\psi b}\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-(\mu+\psi a\sigma^2))^2 - \psi^2 a^2\sigma^4}{2\sigma^2}}\,\mathrm{d}x \\
&\propto \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-(\mu+\psi a\sigma^2))^2}{2\sigma^2}}\,\mathrm{d}x = \mathcal{N}(\mu + a\psi\sigma^2, \sigma^2).
\end{aligned}$$

$\square$

Proposition 4

Proof. For any $x \in \mathbb{R}$, $\pi_f(X = x|E) = \int \pi_f(X = x, Y = y|E)\,\mathrm{d}y \propto \int p(X = x, Y = y|E)e^{\psi f(x,y)}\,\mathrm{d}y = p(X = x)\int p(Y = y|E)e^{\psi f(x,y)}\,\mathrm{d}y$. Hence, for any two values $x_H, x_L \in \mathbb{R}$,

$$\frac{\pi_f(X = x_H|E)}{\pi_f(X = x_L|E)} = \frac{p(X = x_H)}{p(X = x_L)}\frac{\int p(Y = y|E)e^{\psi f(x_H,y)}\,\mathrm{d}y}{\int p(Y = y|E)e^{\psi f(x_L,y)}\,\mathrm{d}y}. \tag{25}$$

Similarly, $\pi_f(X = x) \propto p(X = x)\int p(Y = y')e^{\psi f(x,y')}\,\mathrm{d}y'$, and so

$$\frac{\pi_f(X = x_H)}{\pi_f(X = x_L)} = \frac{p(X = x_H)}{p(X = x_L)}\frac{\int p(Y = y')e^{\psi f(x_H,y')}\,\mathrm{d}y'}{\int p(Y = y')e^{\psi f(x_L,y')}\,\mathrm{d}y'}. \tag{26}$$

$\Pi_{X|E} \succeq_{LR} \Pi_X$ ($\Pi_X \succeq_{LR} \Pi_{X|E}$) if and only if whenever $x_H \geq x_L$ the expression in Equation 25 is greater than or equal (smaller than or equal) than the expression in Equation 26. Equivalently, if the following expression is weakly positive (weakly negative):

$$\frac{\int p(Y = y|E)e^{\psi f(x_H,y)}\,\mathrm{d}y}{\int p(Y = y|E)e^{\psi f(x_L,y)}\,\mathrm{d}y} - \frac{\int p(Y = y')e^{\psi f(x_H,y')}\,\mathrm{d}y'}{\int p(Y = y')e^{\psi f(x_L,y')}\,\mathrm{d}y'}. \tag{27}$$

Moreover, the stochastic dominance is strict if this expression is strictly positive (negative). The expression in Equation 27 has the same sign as

$$\int_y\int_{y'} p(Y = y|E)p(Y = y')\Big(e^{\psi\left(f(x_H,y)+f(x_L,y')\right)} - e^{\psi\left(f(x_H,y')+f(x_L,y)\right)}\Big)\,\mathrm{d}y'\,\mathrm{d}y. \tag{28}$$

Combining terms in which $y < y'$ with terms in which $y > y'$, the expression in Equation 28 equals the following:

$$\begin{aligned}\int_y\int_{y'<y} &\big(p(Y = y|E)p(Y = y') - p(Y = y'|E)p(Y = y)\big)\\ &\Big(e^{\psi\left(f(x_H,y)+f(x_L,y')\right)} - e^{\psi\left(f(x_H,y')+f(x_L,y)\right)}\Big)\,\mathrm{d}y'\,\mathrm{d}y.\end{aligned} \tag{29}$$

In this expression the first term is (strictly) positive if $p(E|Y = y)$ is (strictly) increasing in y, and y is not a.e. constant. Since $y > y'$ and log is a strictly increasing function, the second term has (strictly) the same sign as ψ if f is (strictly) supermodular, and (strictly) the opposite sign if f is (strictly) submodular. The claim follows by combining these observations. $\square$

References

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

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., Wang, X., & Loewenstein, G. (1996). Choosing the wrong pond: Social comparisons in negotiations that reflect a self-serving bias. *Quarterly Journal of Economics*, *111*(1), 1.

Becker, G. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, *76*(2), 169–217.

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

Bracha, A. & Brown, D. J. (2010). Affective Decision Making: A Theory of Optimism Bias. FRB Boston Working Papers Series, paper no. 10-16.

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

Camerer, C. & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, *89*(1), 306–318.

Caplin, A. & Leahy, J. (2001). Psychological Expected Utility Theory and Anticipatory Feelings. *Quarterly Journal of Economics*, *116*(1), 55–80.

Carrillo, J. & Mariotti, T. (2000). Strategic Ignorance as a Self-Disciplining Device. *Review of Economic Studies*, *67*(3), 529–544.

Compte, O. & Postlewaite, A. (2004). Confidence-Enhanced Performance. *American Economic Review*, *94*(5), 1536–1557.

Cowgill, B., Wolfers, J., Wharton, U., & Zitzewitz, E. (2009). Using Prediction Markets to Track Information Flows: Evidence from Google. Mimeo.

De Meza, D. & Southey, C. (1996). The Borrower's Curse: Optimism, Finance and Entrepreneurship. *Economic Journal*, 375–386.

Dillenberger, D., Postlewaite, A., & Rozen, K. (2011). Optimism and Pessimism with Expected Utility. Mimeo.

Dubra, J. (2004). Optimism and overconfidence in search. *Review of Economic Dynamics*, *7*(1), 198–218.

Durlauf, S. & Nagin, D. (2011). Imprisonment and crime. *Criminology & Public Policy*, *10*(1), 13–54.

Edmans, A., García, D., & Norli, Ø. (2007). Sports Sentiment and Stock Returns. *Journal of Finance*, *62*(4), 1967–1998.

Eliaz, K. & Spiegler, R. (2008). Consumer optimism and price discrimination. *Theoretical Economics*, *3*(4), 459–497.

Gottlieb, D. (2010). Imperfect Memory and Choice Under Risk. Mimeo.

Grogger, J. (1991). Certainty vs. Severity of Punishment. *Economic Inquiry*, *29*(2), 297–309.

Heaton, J. (2002). Managerial Optimism and Corporate Finance. *Financial Management*, *31*(2), 33–45.

Hey, J. (1984). The economics of optimism and pessimism. *Kyklos*, *37*(2), 181–205.

Hirshleifer, D. & Shumway, T. (2003). Good Day Sunshine: Stock Returns and the Weather. *Journal of Finance*, *58*(3).

Hoffman, M. (2011a). Learning, Persistent Overconfidence, and the Impact of Training Contracts. Mimeo.

Hoffman, M. (2011b). Overconfidence at Work and the Evolution of Beliefs: Evidence from a Field Experiment. Mimeo.

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. (2006). Ego Utility, Overconfidence, and Task Choice. *Journal of the European Economic Association*, *4*(4), 673–707.

Kőszegi, B. & Rabin, M. (2006). A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics*, *121*(4), 1133–1165.

Landier, A. & Thesmar, D. (2009). Financial contracting with optimistic entrepreneurs. *Review of Financial Studies*, *22*(1), 117–150.

Malmendier, U. & Tate, G. (2008). Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics*, *89*(1), 20–43.

Manove, M. & Padilla, A. (1999). Banking (conservatively) with optimists. *The RAND Journal of Economics*, 324–350.

Mayraz, G. (2011). *Wishful Thinking* (CEP Discussion Paper No. 1092). Centre for Economic Performance, London School of Economics.

Mijović-Prelec, D. & Prelec, D. (2010). Self-deception as self-signalling: a model and experimental evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1538), 227.

Mullainathan, S. & Washington, E. (2009). Sticking With your Vote: Cognitive Dissonance and Political Attitudes. *American Economic Journal: Applied Economics*, *1*(1), 86–111.

Nagin, D. S. & Pogarsky, G. (2001). Integrating celerity, impulsivity, and extralegal sanction threats into a model of general deterrence: Theory and evidence. *Criminology*, *39*(4), 865–892.

Park, Y. & Santos-Pinto, L. (2010). Overconfidence in tournaments: evidence from the field. *Theory and Decision*, *69*(1), 143–166.

Sandroni, A. & Squintani, F. (2007). Overconfidence, insurance, and paternalism. *American Economic Review*, *97*(5), 1994–2004.

Santos-Pinto, L. (2008). Positive Self-image and Incentives in Organisations. *Economic Journal*, *118*(531), 1315–1332.

Saunders, E. J. (1993). Stock Prices and Wall Street Weather. *American Economic Review*, *83*(5), 1337–1345.

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

Sjöberg, L. (2000). Factors in risk perception. *Risk Analysis*, *20*(1), 1–11.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers. *Acta Psychologica*, *47*(2), 143–148.

Washington, E. (2008). Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues. *American Economic Review*, *98*(1), 311–332.

Weinstein, N. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, *39*(5), 806–820.