Priors and Desires

A Model of Payoff-Dependent Beliefs*

Guy Mayraz†

Job-Market Paper

November 25, 2008

Abstract

This paper introduces a model of Bayesian decision making where a person's beliefs about the likelihood of different outcomes depend upon the anticipated payoff consequences of those outcomes. Optimists (pessimists) are more (less) likely to believe A relative to B if the payoff consequences of A are better. Based on the assumption that adding a constant to payoffs does not change the distortion in beliefs, I characterise the unique representation of payoff-dependent beliefs. Choices determine the payoff consequences of outcomes and hence affect the beliefs that in turn determine the optimality of those very choices; choice is thus an equilibrium phenomenon. Optimists may have multiple choice equilibria, and pessimists none. Choice fixes the mapping from states to payoffs, but the payoff consequences of events may remain uncertain. News that affects the expected payoff consequences of an event may therefore alter beliefs, even when it provides no relevant information about its likelihood. Economic consequences are explored in various settings, including the economics of crime, where increasing punishment may encourage, rather than deter crime.

JEL classification: D01,D03,D80,D81,D83,D84.

Keywords: Optimism, pessimism, subjective beliefs, decision under uncertainty, expectations.

---

*I am grateful to Erik Eyster and Richard Layard for their continual guidance and support, to Wolfgang Pesendorfer and Matthew Rabin for stimulating discussions and valuable suggestions at key points in the writing of the paper, and for helpful comments to Michele Piccione and other participants of the LSE Microeconomics Theory Seminar, and to Richard Bradley and other participants of the London-Amsterdam-Lausanne Graduate Workshop in Decisions, Games, and Logic.

†Centre for Economic Performance, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. http://personal.lse.ac.uk/mayraz. g.mayraz@lse.ac.uk.

1 Introduction

Beliefs depend not only on what people *know* to be true, but also on what they *want* to be true. In standard models of subjective beliefs the only way we can explain what a person believes is by reference to the information that person has been exposed to. Any differences in beliefs that cannot be accounted for by asymmetric information can only be modelled as random variation. In this paper I explore the idea that there is a second systematic factor affecting beliefs—that whether the person believes something to be true has to do not only with the information the person possesses but also to do with what the person stands to *gain or lose* from it being true.

The key assumption I make is that there is a time-invariant relationship between people's actual beliefs and the beliefs they would hold if their beliefs depended only on information. The central element in the model is therefore a *distortion mapping* relating the two sets of beliefs as a function of the payoffs the agent faces. I define for such mappings necessary and sufficient properties for a practical model of beliefs that is capable of explaining the key empirical findings. The assumptions I make enable me to derive the following simple representation:

$$\ln \frac{\pi_f(B)}{\pi_f(A)} = \ln \frac{p(B)}{p(A)} + \psi(b-a) \tag{1}$$

where a and b are the payoffs in events A and B respectively, $p(\cdot)$ is the undistorted probability measure, and $\pi_f(\cdot)$ the probability measure distorted by the payoffs defined by f. Beliefs are represented by the log odds ratio between two events A and B sufficiently fine-grained that their respective payoffs are well-defined. According to the equation, the distorted log odds ratio equals the sum of the undistorted log odds ratio plus the parameter ψ times the payoff difference between the two events. The payoff in other events does not enter the equation[1]. Note that the better the agent's information is, the larger $\ln p(B)/p(A)$ is in absolute terms. The parameter ψ therefore plays the role of a *gain factor*, determining both the *direction* that payoffs affect beliefs and the *importance* they have relative to information. If $\psi = 0$ there is no distortion, if $\psi > 0$ the agent's beliefs are distorted in the direction of the event that makes the agent better off, and if $\psi < 0$ the agent's beliefs are distorted in the direction of the event that makes the agent *worse* off.

Equation 1 therefore captures optimism and pessimism. Optimism is identified with $\psi > 0$ and pessimism with $\psi < 0$. Moreover, *relative* optimism (pessimism) is identified with a more positive (negative) value of ψ. The parameter ψ is defined relative to the

[1] This is similar to Bayesian updating of the odds ratio between a pair of events, when only the likelihood odds ratio over that pair of events is required to calculate the posterior odds ratio.

representation of payoffs—if payoffs are scaled, ψ has to be scaled in the inverse direction in order to represent the same distortion. No adjustment to ψ is required if payoffs are shifted by some additive constant, as the payoff differences $b - a$ remain unchanged.

Equation 1 defines a time invariant relationship between distorted beliefs, undistorted beliefs, and payoffs. It follows that, holding undistorted beliefs constant, *changes* in payoffs lead to *changes* in beliefs. This has two key implications:

1. Choice becomes an equilibrium phenomenon. The choice an agent makes determines what payoffs are received in each state. Choice is therefore a function of what states the agent perceives as more likely. But, in the world of Equation 1, beliefs are a function of those same payoffs. Imagine the decision whether to go on a picnic or stay at home. The payoff depends on the weather when picnic is chosen but not when the choice is staying at home. In these conditions Equation 1 implies a distortion of beliefs about the weather if and only if picnic is chosen. An optimist choosing picnic would therefore become more optimistic that the weather would be conducive to a high payoff picnic, making the choice of picnic *self-enforcing* for the optimist. By contrast, a pessimist choosing picnic would become more *pessimistic* about the likelihood of good weather, making the choice of picnic *self-defeating* for the pessimist. More generally, it is possible for optimists to have *multiple* choices that are strictly optimal given the associated beliefs, and for pessimists to have *no* optimal choices from a finite choice set.

2. News can affect beliefs not only through its normative information value but also indirectly, by mediating the effect of belief distortion. In particular, news can affect beliefs even if it is normatively irrelevant. To continue the picnic example, suppose the picnic plans depend upon some other event that has nothing to do with the weather, such as a friend recovering from an illness. Then the friend's recovery is the determinant of expected payoff in different weather conditions, and by Equation 1 news about this event affects beliefs about the weather, even though it provides absolutely no normatively relevant information about it.

The psychology evidence for optimism and pessimism is extensive. In a classical study of optimistic bias, Weinstein (1980) had students rate the relative likelihood of various events happening to them, compared to the likelihood of the same event happening to other students. This difference was interpreted as a measure of the students' optimistic bias[2]. The

[2] Focusing on *comparative* likelihood judgements makes it possible to net out the effects of other belief biases (Sjöberg, 2000).

finding is that students judge desirable (undesirable) events to be more (less) likely to happen to them than to other students. Moreover, the size of the bias correlates with how desirable or undesirable an event is.

Some of the best evidence for payoff-dependent belief distortion comes from experiments in which preferences were exogenously manipulated. In Klein and Kunda (1992) subjects had to assess the ability of a player in a history trivia game. Subjects who were told the person would play on their team (and so wanted him to be a good player) had considerably higher ratings of the player's ability than subjects who expected the person to play on the opposing team (and so wanted him to be a weak player). Changes in payoffs are also central to cognitive dissonance (Cooper and Fazio, 1984; Kunda, 1990).

These and similar studies suggest that what people want affects their beliefs, and that stronger desires affect beliefs more strongly. Moreover, it is clear that the prevailing bias in the population is optimistic. Nevertheless, pessimism is also an important phenomenon, in particular in individuals suffering from depression (Seligman, 1998). Interestingly, one of the symptoms of depression is difficulty making decisions (American Psychiatric Association, 2000), and the purported mechanism seems consistent with the self-defeating dynamics of pessimistic choice as described in the picnic example: "The patient anticipates making the wrong decision: Each time he considers one of the various possibilities he tends to regard it as wrong, and to think that he would regret making the choice." (Beck, 1967).

Payoff-dependent beliefs have been used to analyse a range of economic phenomena, such as bargaining (Babcock and Loewenstein, 1997; Yildiz, 2004), monopolistic contract design (Spiegler and Eliaz, 2008), failure rates and credit rationing in small-business borrowing (De Meza and Southey, 1996), and capital structure in corporate finance (Heaton, 2002). These and other papers have produced many interesting results, demonstrating the range of economic phenomena in which payoff-dependent belief distortion plays an important role.

Akerlof and Dickens (1982) was the first economics model to systematically address belief distortion. More recently, Brunnermeier and Parker (2005) offered a general model of optimism. These and a number of other papers all take a similar approach: they postulate the existence of anticipatory preferences over beliefs, and derive belief distortion as an optimal trade-off between the consumption of desirable beliefs and the costs of belief distortion in terms of misleading future selves into making poor choices. The model of this paper is very different. Perhaps most importantly, a consistent belief distortion applies at all time points given the agent's current information about payoffs. Consider a person deciding whether to get a tent for a wedding reception. In the model of this paper beliefs after hiring the tent would be simply a function of the payoff consequences of different weather conditions. By contrast, in the model of Brunnermeier and Parker (2005) what matters is what decisions are to be made, and very different predictions apply if beliefs are chosen before or after the tent hiring decision. If beliefs are determined *prior* to hiring the tent then there are substantial costs to belief distortion, and limited or no bias may be chosen. If, instead, beliefs are determined *after* the tent has been purchased, then there are *no* costs to belief distortion, and a *total* belief distortion may result.

Beliefs in the model of this paper have nothing to do with the timing of decisions. The important comparative statics have to do instead with information on the one hand, and payoff differences on the other: the higher the payoff differences the greater the bias, and the more information the smaller the bias. These comparative statics are supported by the available evidence (Kunda, 1990; Weinstein, 1980; Sjöberg, 2000). A final important difference is that the model of this paper also extends to pessimism, whereas models in which beliefs are the outcome of choice are necessarily limited to optimism only[3].

Section 2 describes the model. I make a number of assumptions in order to derive a practical formal representation of the mapping relating distorted and undistorted beliefs that captures the essence of payoff-dependent beliefs. First, I assume that the distortion does not change zero probabilities into positive ones, nor positive probabilities into zeroes. Second I assume the mapping is invariant to a relabelling of states. This captures the idea that belief distortion derives from payoffs, rather than with what different states represent. Third, I assume that distorted beliefs conditional on an event depend only on the payoffs in states consistent with that event. Fourth, I assume the distortion is invariant to shifting all payoffs by a constant. This property follows if we identify the agent's *gain* from the event being B rather than A with the *payoff difference* between B and A. Finally, I assume that distorted beliefs are continuous in payoffs. The shift invariance assumption is perhaps more controversial than the others. It is therefore noteworthy that a much weaker condition is sufficient to derive a close variant of Equation 1 in which the $\psi(b-a)$ term is replaced by $\psi(v(b)-v(a))$ for a monotonically increasing function v.

Equation 1 has the same form as Bayes Rule[4]. Taking this interpretation, the effect of payoff on beliefs can be seen notionally as *evidence*, with optimists, for example, assigning higher likelihood to high-payoff states. Choice aside, agents can therefore be seen as standard

[3]It is not implausible that some people would choose to have positively biased beliefs, but it *is* implausible that pessimists *would choose* to have negatively biased beliefs.

[4]In log odds terms Bayes Rule is

$$\frac{\pi(B|I)}{\pi(A|I)} = \frac{\pi(B)}{\pi(A)} + \frac{\pi(I|B)}{\pi(I|A)} \tag{2}$$

Undistorted beliefs take the role of the prior odds ratio, ψ times the payoff difference is the evidence, and the distorted odds ratio is the posterior.

Bayesian updaters except for the systematic dependence of the prior on payoffs. One important implication is that beliefs retain the martingale property that agents do not expect to revise their expectations on average after receiving new information.

In Equation 1 good information translates into a big (in absolute terms) $\log p(B)/p(A)$ term, and hence a relatively small bias in probability terms, while a big payoff difference translates into a relatively large bias. More generally the greater the stakes the a person has in an outcome the greater the bias, and the more information the smaller the bias. These comparative statics are seen most clearly if payoffs are some factor times a normally distributed variable. The distortion then maintains the normal distribution but shifts the mean as a function of the size of the stakes (represented by the factor relating payoff to the normal variable) and of the amount of information (as measured by the variance).

Section 3 explores choice behaviour. Equation 1 associates with each act f a distorted probability measure π_f, but it doesn't say what determines the reference act f. To close the model I assume that resolving to choose an act f makes f into the reference act. f is then a *rational choice* if it maximises expected utility according to π_f, the distorted beliefs that the agent has when f is the reference act. More generally, the beliefs π_f induce a complete *act-specific preference relation* $\succeq_f$ over the set of acts. f is thus rational if and only if $f \succeq_f g$ for all acts g in the choice set. This notion of rational choice can also be thought of as a *choice equilibrium*. This term is particularly appropriate if we think of the choice process in dynamic terms. Contemplating the choice of f shifts beliefs towards π_f, and may either push preferences further in the direction of f, or else pull the decision maker away from f. The stable points of this dynamics correspond to choice equilibria as defined above.

Imagine a person has to decide whether to go on a picnic or stay at home, and suppose picnic beats staying at home if and only if the weather is nice. Then it is possible that for an optimist (picnic $\succ_{\text{picnic}}$ home) and (home $\succ_{\text{home}}$ picnic), so that *both* picnic and staying at home are rational choices, and that for a pessimist (picnic $\prec_{\text{picnic}}$ home) and (home $\prec_{\text{home}}$ picnic), so that *neither* option is rational. More generally, consider any finite choice set that consists of independent acts (the payoff of which is a function of independent random variables), such that the expected payoff from each act lies strictly within the possible payoff distribution of the other acts. Then a sufficiently optimistic agent has multiple rational choices and that a sufficiently pessimistic agent has none.

Suppose an agent can choose any mixing of payoffs from a finite choice set. By a standard fixed-point argument there is then at least one choice equilibrium. In game theory each act in the support of a mixed equilibrium is a best response. An agent faced with a finite choice set can thus play a mixed equilibrium action by randomising, and then choosing an act from the original choice set as a function of the outcome of a randomising device. This, however, is *not* the case with pessimistic choice, as the acts in the support of a pessimistic mixed equilibrium are generally not optimal. The pessimist will therefore be unable to rationally follow through on the choice of act indicated by the result of the randomising device. A mixed equilibrium is therefore only a realistic option for a pessimist if the pessimist can *irreversibly commit* to the mixing. The practical utility of the theoretical existence of mixed choice equilibria for pessimists is therefore limited.

That an optimist has an optimal choice from any finite choice set is trivial in the case of independent acts, but is otherwise not so obvious. As a step in proving this result I prove that the stability of strictly mixed equilibria is determined by whether or not an agent is optimistic or pessimistic. In particular, the strictly mixed equilibria of optimists are always unstable. Using this result it is easy to show that an optimist must have a rational choice from the original finite choice set.

The fact that optimists may have multiple choice equilibria has similarities to the personal equilibrium predictions of Kőszegi and Rabin (2006, 2007) for agents with loss-averse preferences. There are notable differences, however. For instance, loss-aversion never results in the choice of dominated alternatives, whereas it can be a choice equilibrium for an optimistic agent to choose a strictly dominated option, such as an investment associated with a first-order stochastically dominated payoff distribution.

In Section 4 I look at how the subjective probability of an event can change as a result of news about its payoff consequences, even if the news provides no relevant information about its likelihood. News can thus affect beliefs not only directly (by providing relevant evidence), but also *indirectly*, by mediating the effects of belief distortion. The key proposition in this section shows that whenever two variables are complements in the agent's utility function, good news about one variable increases the bias in beliefs about the second variable. The case of substitutes goes in the other direction.

Consider again the picnic example, but suppose now that the picnic plans are contingent on a friend recovering from an illness. This has the effect of letting an *outside event*—whether the friend recovers in time—determine the payoff in the events of rain or shine. As in the analysis of Section 3 the belief bias depends on whether or not the picnic is on. It follows that if the friend remains ill the subjective probability of rain would go up, and if the friend recovers it would go down. This, of course, is in spite the fact that the friend's recovery from illness provides no normatively relevant information about the weather.

The key to this example is that going out on a picnic and a good weather are *complements*. A predictable non-normative change in beliefs also occurs in the case of *substitutes*. For example, suppose a sales person needs to hit a certain target in order to be promoted, and is one deal short of this target. Then different prospective sales that could help achieve the target are substitutes. The prediction is that success in one such prospect (thereby hitting the target) reduces the bias on the others, whereas a failure leads to an increase in the bias over any remaining prospects.

On the assumption that must people are optimistic payoff-dependent belief update accounts for some of the key findings in experiments exploring motivated cognition and cognitive dissonance. For example, Klein and Kunda (1992) found that the expectation to play with or against a given player alters beliefs about the ability of that player. Letting x denote the ability of the player, and $a \in \{-1, 1\}$ the side on which she will play, the agent's payoff can be modelled by $u(a, x) = ax$. Thus, a and x are *complements*, and the prediction of the model agrees with the empirical findings.

Section 5 considers choices between risky and safe alternatives. Optimists overestimate gains and underestimate risks, and pessimists do the reverse. Consequently, controlling for the utility function, optimists make relatively risk-loving choices, whereas pessimists are risk-averse. Optimistic/pessimistic revealed risk preferences can be distinguished from underlying risk preferences by comparing bets in conditions of subjective uncertainty with bets in conditions of objective uncertainty.

An increase in the gap between good and bad outcomes leads to an increase in the bias. Let f and f' be binary bets over events A and B such that $f(B) = f'(B) > f'(A) > f(A)$. Then by Equation 1, $\pi_f(B)/\pi_f(A) > \pi_{f'}(B)/\pi_{f'}(A)$, and it can even be possible that $\mathcal{E}_f(f) > \mathcal{E}_{f'}(f')$ even though f' first-order stochastically dominates f. In an example from the economics of crime I show that an increase in punishment contingent on getting caught can lead to an *increase* in subjective expected utility, and hence to a *decrease* in deterrence. Finally, I show that while first-order stochastic dominance can be violated, stochastic dominance in the likelihood ratio is preserved under payoff-dependent belief distortion. In particular, for f' to stochastically dominate f in the likelihood ratio is a sufficient condition for $\mathcal{E}_{f'}(f') \geq \mathcal{E}_f(f)$.

2 Belief distortion

2.1 Model

Let $S = [0, 1]$ denote the state space and $F = \{f : S \to \mathbb{R}\}$ the collection of simple Anscombe-Aumann acts (for a simple act f, the set of its prizes $I_f = \{a \in \mathbb{R} : \exists s \in S, f(s) = a\}$ is finite). The prize in state s should be interpreted as the expected utility of an appropriate objective lottery. Let Σ denote the Borel σ-algebra on S, and let Δ denote the set of probability measures on Σ. The probability measure $\mu \in \Delta$ is *non-atomic* if for every event A and $\lambda \in [0,1]$ there is an event $B \subseteq A$ with $\mu(B) = \lambda\mu(A)$. For $\mu \in \Delta$ and an event A with $\mu(A) > 0$ let $\mu(\cdot|A)$ denote the conditional probability measure on A. I write a, b for constant acts (that yield utility a, b in every state) and f, g, h for generic acts. I write aAg for the act that yields a on the event A, and g on A^c. For disjoint events A and B I write $aAbBh$ for an act that yields a on A, b on B, and h on $(A \cup B)^c$. For an act f and a prize $a \in I_f$ I write $f_{(a,b)}$ for an act that yields b for states in which f yields a, and is otherwise identical to f.

The primitives of the model are (i) a non-atomic probability measure p (the undistorted measure), and (ii) a map π that associates to each simple act $f \in F$ a distorted probability measure π_f. The first definition describes the basic properties we want this distortion to satisfy, and the second definition describes the logit formula. The theorem says that the two definitions are equivalent.

Definition 1 (Payoff-Dependent Distortion)**.** A map $\pi : F \to \Delta$ is a *payoff-dependent distortion* of p if the following properties are satisfied:

A1 (state independence). If $p(A) = p(B)$ for disjoint events A and B, and $f = aAbBh, g = bAaBh$, then $\pi_f(A) = \pi_g(B)$.

A2 (information independence). If $g(s) = f(s)$ for all $s \in A$, then $\pi_f(\cdot|A) = \pi_g(\cdot|A)$.

A3 (absolute continuity). $\pi_f(A) = 0 \iff p(A) = 0$ for all events A.

A4 (prize continuity). Suppose $a_n \to a$ and let $f_n = f_{(a,a_n)}$ then $\pi_{f_n}(\cdot) \to \pi_f(\cdot)$.

A5 (shift invariance). If $f(s) = g(s) + a$, then $\pi_f = \pi_g$.

A1 requires that the labelling of states plays no role in the distortion mapping. This assumption ensures that the distortion is a function only of payoffs. A2 requires that distorted beliefs conditional on an event depend only on the payoffs in states that are consistent with that event. This assumption has the flavour of an independence of irrelevant alternatives assumption, and has a rough analogue in Luce's Choice Axiom (Luce, 1959)[5]. A3 says that the log odds ratio between two events may be shifted as the result of the distortion, but only by a finite amount. A4 imposes continuity in prizes, so that similar acts result in a similar

[5]Luce's Choice Axiom is defined in the context of a theory of probabilistic choice. It requires that the odds ratio for choosing x over y be independent of what other alternatives are in the choice set.

distortion. To understand assumption A5 note that the distortion mapping is intended to capture the idea that whether the agent believes A or B depends on what the agent has to gain or lose if B rather than A were the case. A5 identifies the agent's *net gain* from the event being B rather than A with the *payoff difference* between B and A. This assumption is perhaps more controversial than the others. It is therefore noteworthy that a much weaker condition is sufficient to derive a close variant of Equation 3 in which the $\psi(b-a)$ term is replaced by $\psi(v(b)-v(a))$ for a monotonically increasing function v[6].

Definition 2 (Logit Distortion)**.** A map $\pi : F \to \Delta$ is a *logit-distortion* of p if the following two conditions are satisfied:

1. $\pi_f(A) = 0 \iff p(A) = 0$ for all events A.
2. There exists $\psi \in \mathbb{R}$ such that for all prizes $a, b \in I_f$ and positive probability events $A \subseteq f^{-1}(a)$ and $B \subseteq f^{-1}(b)$,

$$\ln \frac{\pi_f(B)}{\pi_f(A)} = \ln \frac{p(B)}{p(A)} + \psi(b-a) \tag{3}$$

In Definition 2 the distortion is defined in terms of its effect on the log odds between pairs of payoff-specific events. Since any event can be written as the disjoint union of such events, this definition is sufficient to fully specify the distorted probability measure π_f. Further remarks about Equation 3 are deferred to Section 2.2.

Proposition 1 (Representation theorem)**.** *A map $\pi : F \to \Delta$ is a logit distortion of p if and only if it is a payoff-dependent distortion of p.*

The proof follows three lemmas. The first establishes that the agent does not distort the relative probability of states over which he is indifferent. The second shows that the distortion of the relative probabilities of three equi-probable events with three prizes a, b and c satisfying $c - b = b - a$ is the same, namely $\pi_f(c)/\pi_f(b) = \pi_f(b)/\pi_f(a)$. The third lemma uses these results to relate the distortion of the relative probabilities between any two pair of events to the ratio of the difference in prizes in each of the two pairs. Finally, in the main

[6]The essence of the proof that such a representation exists is that for all events A, B and X, $\log \pi_f(B)/\pi_f(A) = \log \pi_f(B)/\pi_f(X) - \log \pi_f(A)/\pi_f(X)$. The linearity of this expression makes it possible to define v consistently for all events. Using induction on the set of prizes in a simple act it is possible to construct a *single* representation of this type that works for *all* simple acts with rational prizes. Continuity of prizes then ensures that this result extends to all simple acts with real prizes. Finally, given the prize continuity assumption, a requirement that for all acts f and prizes a and b if $p(f^{-1}(a)) = p(f^{-1}(b))$ and $\pi_f(f^{-1}(a)) = \pi_f(f^{-1}(b))$ then $a = b$ is a sufficient condition ensure that v is monotonic.

proof a pair of reference events is chosen, ψ is defined to solve Equation 3 for the reference events, and the lemmas are used to prove that the same ψ represents the distortion for all pairs of events.

The intuition for this result is that (1) by absolute continuity, the distortion can be described by a likelihood function; (2) by state independence, this function depends on states only via the payoffs in these states; (3) by information independence, the distortion over pairs of events can depend only on the payoffs in the two events; (4) by shift invariance, only the difference in payoffs matter; (5) for any three events A, B and C, $\ln \pi_f(B)/\pi_f(A) = \ln \pi_f(B)/\pi_f(C) + \pi_f(C)/\pi_f(A)$. Together with (4) this makes it possible to relate the distortion of the odds ratio between any two pairs of events, such that the difference in payoffs within each pair are a rational multiple of the difference in payoffs within the other pair, and (6) by continuity in prizes, this result extends to all pair of events.

Lemma 1. *Suppose $B \subseteq A = f^{-1}(a)$ for some $a \in I_f$ and that $p(A) > 0$ then $\pi_f(B)/\pi_f(A) = p(B)/p(A)$.*

Lemma 2. *Suppose $c - b = b - a > 0$ for some $a, b, c \in I_f$, and let $A \subseteq f^{-1}(a)$, $B \subseteq f^{-1}(b)$, $C \subseteq f^{-1}(c)$, such that $p(A) = p(B) = p(C) > 0$. Then,*

$$\frac{\pi_f(C)}{\pi_f(B)} = \frac{\pi_f(B)}{\pi_f(A)} \tag{4}$$

Lemma 3. *Suppose $A = f^{-1}(a)$, $B = f^{-1}(b)$, $A' = f^{-1}(a')$, and $B' = f^{-1}(b')$ are events in the support of p such that $a < b$, $a' < b'$ and $a \leq a'$. Then*

$$\ln\left(\frac{\pi_f(B')}{\pi_f(A')} \cdot \frac{p(A')}{p(B')}\right) = \frac{b' - a'}{b - a} \cdot \ln\left(\frac{\pi_f(B)}{\pi_f(A)} \cdot \frac{p(A)}{p(B)}\right) \tag{5}$$

2.2 How belief distortion works

Equation 3 describes the distorted beliefs of the agent as the joint function of what the agent knows and what the agent wants to be true. The same beliefs can result from different combinations of information and payoffs (Figure 1). The parameter ψ plays the role of a gain factor, determining both the direction that payoffs affect beliefs, and the importance of their role. If $\psi > 0$ the agent's beliefs are distorted in the direction of events that make the agent better off. If, instead, $\psi < 0$ the agent's beliefs are distorted in the direction of events that make the agent *worse* off. Finally, if $\psi = 0$ the agent's beliefs are objective—not in the sense of being necessarily correct, but in the sense that what the agent believes is independent of what the agent wants to be true. ψ is defined relative to the representation

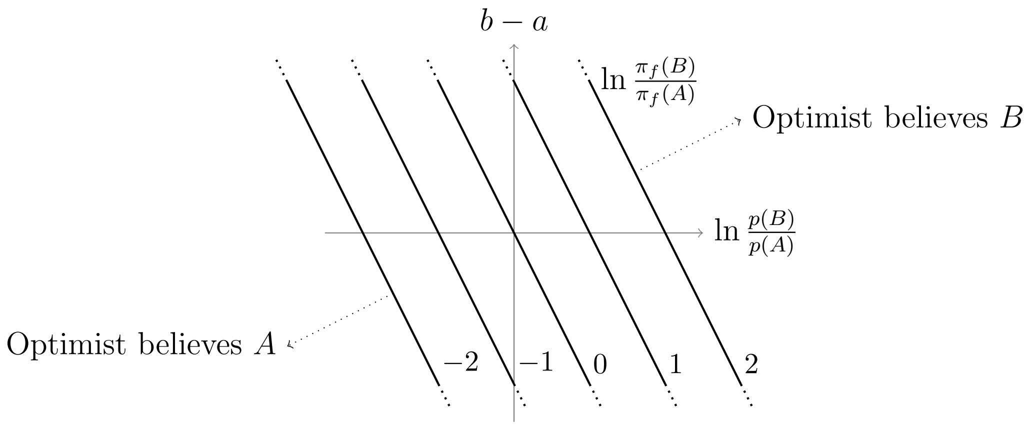

Figure 1: Iso-belief lines as a function of unbiased beliefs the stakes in the event being B rather than A. Unbiased beliefs are plotted on the x-axis and the stakes are plotted on the y-axis. Iso-belief lines are straight with slope $1/\psi$. This figure is plotted for an optimist. Iso-belief lines for a pessimist slope in the opposite direction, while those of a neutral agent are vertical.

of payoffs—if payoffs are scaled, ψ has to be scaled in the inverse direction in order for it to represent the same distortion. No adjustment to ψ is required if payoffs are shifted by some additive constant, as the payoff differences $b - a$ remain unchanged.

Fixing the act f we can define the *likelihood* of the payoffs by $L(f|E) = e^{\psi f(E)}$ where E is any event for which the payoff is well-defined. Using this definition Equation 3 takes the form of Bayesian updating:

$$\ln \frac{\pi_f(B)}{\pi_f(A)} = \ln \frac{p(B)}{p(A)} + \frac{L(b|B)}{L(a|A)} \tag{6}$$

This notional equivalence implies that the martingale property of consistent beliefs is retained, as if agents in the model were standard Bayesian agents. In particular, in order to obtain the agent's distorted posterior beliefs following the arrival of new information, we can *either* (i) apply the Bayesian update to compute the undistorted posterior beliefs, and then use Equation 3, *or* (ii) use Equation 3 to obtain the distorted prior beliefs, and the apply the Bayesian update.

From this perspective there are two key differences between agents in the model and standard agents: (i) the distorted prior is only well-defined after the choice of act, and (ii) the distorted prior depends *systematically* on payoffs, so that differences in payoffs can be used to explain differences in beliefs.

2.3 Belief distortion over payoffs

What belief distortion implies for expected payoff is central to our notion of optimism and pessimism, and is also the key to determining how belief distortion affects preferences between acts. In this sub-section I formally define beliefs over payoffs, expectations over those payoffs, and the bias size. I then consider specifically the case of acts with a normally distributed payoff. The attractive property of such acts is that belief distortion takes a particularly clean form: since the payoff likelihood term is exponential (Equation 6) the distorted distribution remains normal, and the distortion is limited to a shift of the mean of the normal distribution (See also Figure 2).

Definition 3 (Distribution of payoffs from an act)**.** For an act f let X_f denote the real random variable defined by $X_f(s) = f(s)$. The *undistorted probability distribution* of f, $F(x)$, is the probability distribution of X_f using the undistorted probability measure p. That is, $F(x) = p(X_f \leq x)$. For any acts f and g the *distorted probability distribution of f conditional on g*, $F_g(x) = \pi_g(X_f \leq x)$ is the probability distribution of X_f using the distorted probability measure π_g. The *undistorted expectation of f* is $\mathcal{E}(f) = \sum_{s \in S} p(s) f(s)$. The *expectation of f conditional on g* is $\mathcal{E}_g(f) = \sum_{s \in S} \pi_g(s) f(s)$. The *payoff bias* of an act f is the difference between the distorted and undistorted expectations with f as the reference: $\mathcal{E}_f(f) - \mathcal{E}(f)$.

Proposition 2 (Normally distributed payoffs)**.** *Let f be an act such that $F(x)$ is a normal distribution with mean μ and variance σ^2. Then $F_f(x)$ is also a normal distribution with the same variance σ^2, but with mean shifted to $\mu + \sigma^2 \psi$.*

2.4 Comparative statics

Equation 3 can be written qualitatively as follows:

$$\text{Beliefs} = \text{Information} + \psi \cdot \text{Stakes} \tag{7}$$

Beliefs are thus a tug-of-war between information on the one hand, and what the agent has at stake on the other. The implied comparative statics for the belief bias are: (i) the more is at stake, the bigger the bias, and (ii) the stronger the evidence, the *weaker* the bias. This can be seen most clearly for (i) binary payoffs, and (ii) normally distributed payoffs. In the binary case suppose the payoff of an act f is v if some event E obtains and zero otherwise.

Then using Equation 6 the bias in expected utility terms is given by

$$\begin{aligned}\mathcal{E}_f(f) - \mathcal{E}(f) = (\pi_f(E) - p(E))v &= \left(\frac{p(E)e^{\psi v}}{1 - p(E) + p(E)e^{\psi v}} - p(E)\right)v \\ &= \frac{(e^{\psi v} - 1)p(E)(1 - p(E))}{1 + p(E)(e^{\psi v} - 1)}v\end{aligned} \tag{8}$$

The sign of this expression is that of ψ. Its absolute size is increasing in the stakes v, and is decreasing in the limit of strong information ($p(E) \to 0$ or $p(E) \to 1$), in accordance with the comparative statics above.

In the normal case the bias is proportional to the undistorted variance of the payoff (Proposition 2). Suppose there is a normally distributed random variable X such that the payoff of f is proportional to X, i.e. $f = aX$. The variance of f is then equal to a^2/τ where $\tau = 1/\sigma^2$ is the *precision parameter* of X. Information can be identified with τ, and the stakes with a. The bias is thus proportional to the square of the stakes, and inversely proportional to information.

The evidence for the comparative statics of evidence is overwhelming (Kunda, 1990). For the comparative statics for the strength of preferences over events see Weinstein (1980) and Sjöberg (2000). There is also strong evidence that *changes* in stakes can lead to *changes* in beliefs. The theory and evidence for this are explored in Section 4.

It is interesting to note that models of anticipatory preferences (Akerlof and Dickens, 1982; Brunnermeier and Parker, 2005) produce very different comparative statics. The most obvious comparative statics of such models has to do with the length of time in which the agent enjoys anticipatory utility: the longer this time, the stronger bias. Anticipatory utility models have no obvious comparative statics for the stakes the agent has in a given situation. In fact, if anticipatory utility is increasing less than linearly in the stakes the prediction would be for the *opposite* comparative statics. Finally, the comparative statics for information are unclear, and depend on whether a cost of information distortion is introduced. If there are no such costs then there are cases in which anticipatory utility models would predict *total* bias[7].

[7] It makes a big difference for anticipatory utility models whether the decision the agent has to take is continuous or discrete. If the latter, then depending on the optimal trade-off, the agent may either choose just enough bias so as not to make a costly error, or else accept the error and enjoy as much bias as possible.

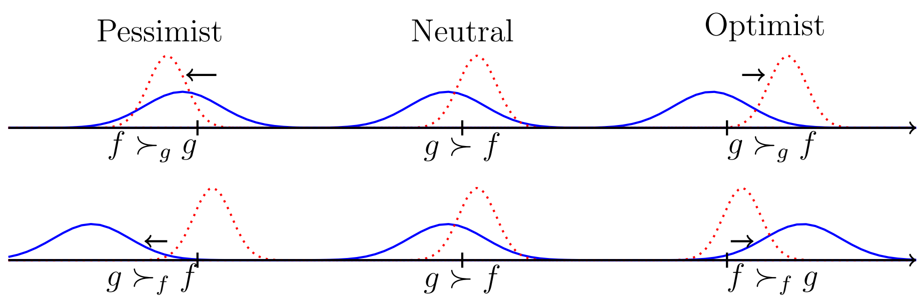

Figure 2: Belief distortion in binary choice between two independent alternatives: f in the solid blue line, and g in the red dotted line. The x-axis denotes expected utility. The top panel shows beliefs with g as reference, and the bottom panel beliefs with f as reference. As drawn, g is the efficient choice, both f and g are rational choices for the optimistic agent, and neither is a rational choice for the pessimistic agent. Comparing f and g, note that there is more disparity in f, and hence the distortion shifts f more than it shifts g (Section 2.4).

2.5 Optimism and pessimism

Equation 3 suggests that a positive value for ψ can be identified with optimism, and a negative value with pessimism. In this section I consider how the distribution of payoffs is distorted under different values of ψ. The main result is that the distorted distributions corresponding to different values of ψ are related to each other via a monotonic likelihood ratio. In particular, it follows that, whatever the act, agents with a higher value of ψ expect a higher payoff. ψ can thus be used as a model of *relative* as well as *absolute* optimism and pessimism.

Stochastic dominance in the likelihood ratio is defined for distributions with the same support. This, of course, is the case for the distribution of payoffs for the same act distorted by different values of ψ.

Definition 4. Suppose $F(x)$ and $G(x)$ are two probability distributions with the same support then $F(x)$ *stochastically dominates* $G(x)$ *in the likelihood ratio*, written $F \succeq_{LR} G$ if $dF(x)/dG(x)$ is non-decreasing for all x in the support of F and G.

Proposition 3 (Relative optimism)**.** *Let f be any act, and let $F_f^{\psi}(\cdot)$ denote the distorted distribution of the payoff of f with act f as the reference for an agent with a distortion parameter ψ. Then $F_f^{\psi} \succeq_{LR} F_f^{\psi'}$ if $\psi \geq \psi'$.*

The proof is immediate from Equation 1. Note that the claim is for the payoff expected from act f when f is the reference. The effect of ψ on the expected payoff from some other act g depends on how the two acts are related to each other. If, for example, $g = -f$ then the effect of ψ on the distorted distribution of g will be the opposite of its effect on the distorted distribution of f.

It is also interesting to consider the limit cases of $\psi \to \pm\infty$. In the limit of $\psi \to \infty$ *all* the probability is put on the best possible outcome. That is, an extremely optimistic agent believes in the best possible world for him. Similarly, an extreme pessimist believes in the *worst* possible world. Note that extreme optimism and pessimism are not themselves logit distortions (they violate absolute continuity, as events resulting in other than the extreme payoff are distorted to zero probability). Yildiz (2007) explores a model in which agents behave like extreme optimists in this sense.

Proposition 4 (Extreme optimism/pessimism)**.** *Let f be a simple act. Define $a_{min} = \min_{s \in Supp(p)} f(s)$ and $a_{max} = \max_{s \in Supp(p)} f(s)$ to be the minimal and maximal possible prizes respectively, and let $A_{min} = f^{-1}(a_{min})$ and $A_{max} = f^{-1}(a_{max})$ denote the events that these prizes are obtained. Then $\lim_{\psi \to -\infty} \pi_f(A_{min}) = \lim_{\psi \to \infty} \pi_f(A_{max}) = 1$.*

3 Choice equilibrium

I start with an example, and then proceed to analyse formally how choice works in the model.

Example 1. Picnic

Suppose a person has to decide whether or to go on a picnic or stay at home, and that the payoff for going out on a picnic is weather dependent, whereas the payoff for staying at home is not:

	Rain	Shine
Picnic	0	1
Home	0.5	0.5

Since the payoff difference between "rain" and "shine" depends on the choice between picnic and staying at home, Equation 3 implies that beliefs over these events also depend on the choice. Suppose, for example, that the undistorted probability for a sunny day is $p(\text{shine}) = 1/3$, and that the agent is optimistic with $e^{\psi} = 4$. With these numbers the distorted probability of a sunny day conditional on the choice of picnic is

$$\pi_{\text{picnic}}(\text{shine}) = \frac{p(\text{shine})e^{\psi}}{p(\text{shine})e^{\psi} + p(\text{rain})} = \frac{(1/3) \cdot 4}{(1/3) \cdot 4 + (2/3) \cdot 1} = 2/3 \tag{9}$$

whereas the distorted probability of the same event conditional on staying at home is

$$\pi_{\text{home}}(\text{shine}) = \frac{p(\text{shine})e^{\psi/2}}{p(\text{shine})e^{\psi/2} + p(\text{rain})e^{\psi/2}} = \frac{(1/3)\cdot 2}{(1/3)\cdot 2 + (2/3)\cdot 2} = 1/3 \tag{10}$$

It follows that if picnic is the reference act, the subjective expected utility of picnic is $\mathcal{E}_{\text{picnic}}(\text{picnic}) = 2/3$, whereas if staying at home is the reference, the subjective expected utility of picnic is only $\mathcal{E}_{\text{home}}(\text{picnic}) = 1/3$. Since the expected utility of staying at home is 1/2 regardless of the reference act, it follows that *both choices are (subjectively) strictly optimal*.

For an example of choice by pessimists, suppose the agent is pessimistic with $e^{\psi} = 1/4$, and that the undistorted probability of a sunny day is $p'(\text{shine}) = 2/3$. Then by a similar calculation $\mathcal{E}'_{\text{picnic}}(\text{picnic}) = 1/3$, $\mathcal{E}'_{\text{home}}(\text{picnic}) = 2/3$, and $\mathcal{E}'_{\text{picnic}}(\text{home}) = \mathcal{E}'_{\text{home}}(\text{home}) = 1/2$, so that *neither choice is subjectively optimal*.

In the following I formally discuss choice, and investigate the conditions for the existence and uniqueness of rational choice.

3.1 Rational choice

The preferences of an Anscombe-Aumann agent over the set F of simple acts correspond to comparisons of subjective expected utility. Act g is preferred to h if and only if the subjective expected utility of g is higher than that of h:

$$g \succeq h \iff \sum_{s\in S} p(s)\{g(s) - h(s)\} \geq 0 \tag{11}$$

The beliefs of an agent with payoff-dependent beliefs depend on the reference act f. Equation 11 continues to hold, but $p(s)$ is replaced by $\pi_f(s)$. As a result, preferences over acts become a function of f:

$$g \succeq_f h \iff \sum_{s\in S} \pi_f(s)\{g(s) - h(s)\} \geq 0 \tag{12}$$

or in a different notation,

$$g \succeq_f h \iff \mathcal{E}_f(g) \geq \mathcal{E}_f(h) \tag{13}$$

Choosing an act f from some choice set $C \subseteq F$ fixes f as the reference act. f is then a *rational choice* if $f \succeq_f g$ for all $g \in C$. A rational choice can also be thought of as a *choice equilibrium*. In the picnic example the optimist has two rational choices (or two choice equilibria), whereas the pessimist has none.

A standard agent may rationally choose both f and g from a choice set, but only if $g \sim f$. By contrast, an optimist can have multiple *strict* equilibria, such that both $f \succ_f g$ and $g \succ_g f$. Obviously, at most one can be a good choice in the sense of maximising expected utility given undistorted beliefs. It follows that optimists may choose inefficient choices in equilibrium.

Pessimists may have *no* rational choice from a finite choice set. Interestingly, one of the symptoms of depression is difficulty making decisions (American Psychiatric Association, 2000), and the underlying reason can be interpreted as having no choice equilibrium: "The patient anticipates making the wrong decision: Each time he considers one of the various possibilities he tends to regard it as wrong, and to think that he would regret making the choice." (Beck, 1967).

In the next section I look at the general case of independent acts. I first show that the choice of one such act does not distort beliefs over other independent acts. I then use this result to provide a weak condition for binary choice sets that ensures a sufficiently optimistic agent has multiple rational choices, and that a sufficiently pessimistic agent has none.

3.2 Independent acts

Intuitively, if alternatives are unrelated, the bias induced by choosing one should not affect beliefs about the others. In the following I show that this is indeed the case. To formalise the notion of unrelated acts I look at the random variables that describe their payoffs, and define the acts as independent if those random variables are independent. Independence is defined relative to the undistorted probability measure[8].

Definition 5 (Independent acts)**.** Acts f and g are *independent* if there exist random variables X_f and X_g such that $f(s) = f(X_f(s))$ and $g(s) = g(X_g(s))$ for all $s \in S$, and X_f and X_g are independent relative to the undistorted probability measure $p(\cdot)$.

Given this definition it follows that having act f as the reference biases beliefs over f only. The proof works by expressing the relevant events for g as the union of events that specify values for both f and g. The likelihood term of the belief distortion (Equation 6) then drops out, as it depends only on f, and f and g are assumed to be independent.

[8] Independence of random variables relative to the undistorted beliefs is, in general, neither a necessary nor a sufficient condition for independence relative to the distorted probability measure. See Section 4.

Proposition 5 (Independent acts)**.** *Suppose f and g are independent acts then the distorted probability distribution of g conditional on f is the same as its undistorted probability distribution: $G_f(x) = G(x)$ for all $x \in \mathbb{R}$.*

Suppose now that f and g are independent acts, and that the agent is extremely optimistic (pessimistic). Then conditional on selecting g the agent perceives g as delivering the best (worst) possible payoff for g. It follows that *both* f and g are rational choices for a sufficiently optimistic agent, unless the expected value of one is better than the *best* possible payoff of the other. Similarly, *neither* f nor g are rational choices for a sufficiently pessimistic agent, unless the expected value of one is worse than the *worst* possible payoff of the other:

Proposition 6. *Suppose f and g are independent acts, such that $\mathcal{E}(f) < \max_{s \in Supp(p)} g(s)$. Then there is ψ^* such that g is a rational choice for any agent with a distortion parameter $\psi \geq \psi^*$. Similarly, if $\mathcal{E}(f) > \min_{s \in Supp(p)} g(s)$ then there is a distortion parameter ψ^*, such that g is not a rational choice for any agent with a distortion parameter $\psi \leq \psi^*$.*

3.3 Mixed equilibria

By a standard fixed-point argument if the choice set C is the mixture space over a final set of acts then the agent has a rational choice. It may seem, therefore, that a pessimist faced with a finite choice set can compute the mixed equilibrium, and emulate it using a randomising device. This intuition is wrong, however, as it would be irrational for the pessimist to follow through on the result of the randomisation. This highlights an important difference between the mixed equilibria of pessimists and the mixed equilibria we normally encounter in Game Theory. In games like Matching Pennies the acts in the support of a strictly mixed equilibrium are themselves optimal choices, but the acts in a mixed equilibrium of a pessimistic agent are generally *not* optimal. The pessimist will therefore be unable to rationally follow through on the choice of act indicated by the result of the randomising device. A mixed equilibrium is therefore only a realistic option for a pessimist if the pessimist can *irreversibly commit* to the mixing[9], but actual choice situations need not offer this possibility even if randomising devices are available.

Another interesting question about equilibria in general and strictly mixed equilibria in particular is their stability. If we think as choices and beliefs as either being in equilibrium or not being in equilibrium, then any fixed point is like any other. But if we think of choices and beliefs in dynamic terms, then it makes sense to ask whether a particular fixed point is

[9]One interesting example of commitment is to hand over the decision to some other person (who then decides by tossing a coin, or else makes a decision which the pessimist is unable to predict).

stable or unstable. In particular, suppose $f = \alpha g + (1-\alpha)h$ is a strictly mixed equilibrium. Then the agent is necessarily indifferent between g and h in the equilibrium f. Suppose however that the agent contemplates a mix of g and h with a somewhat higher proportion $\alpha' > \alpha$ of g. Given this new reference act it is possible that the agent prefers g to h, in which case the deviation will be self-enforcing and the equilibrium *unstable*, and it is also possible that the agent may then prefer h to g, in which case the deviation will be self-defeating and the equilibrium *stable*.

The striking result I show below is that the stability of strictly mixed equilibria is fully determined as a function of whether the agent is an optimist or pessimist. The strictly mixed equilibria of optimists are all unstable, and those of pessimists are all stable. The intuition is that increasing the proportion of some alternative in a strictly mixed equilibrium increases the bias over that alternative and reduces the bias over the other. Since the optimistic bias is positive and the pessimistic bias negative, an optimist will tend to move away from equilibrium, and a pessimistic will move back towards equilibrium. The definition and proofs assume the equilibrium has only two acts in its support, but since a mixed act is itself an act there is no loss of generality[10].

Definition 6 (Stability of strictly mixed equilibria)**.** Suppose $f = \alpha g + (1-\alpha)h$ is a strictly mixed equilibrium for $g, h \in C$, and let $f(\epsilon) = (\alpha+\epsilon)g + (1-(\alpha+\epsilon))h$. If there is $\delta > 0$ such that for all $0 < \epsilon < \delta$, $g \succeq_{f(\epsilon)} h$ then f is an *unstable equilibrium*. Similarly, if for all $0 < \epsilon < \delta$, $g \preceq_{f(\epsilon)} h$ then f is a *stable equilibrium*.

Proposition 7. *If $f = \alpha g + (1-\alpha)h$ for $\alpha \in (0,1)$ is a strictly mixed choice equilibrium then f is a stable if and only if $\psi \leq 0$ and is unstable if and only if $\psi \geq 0$.*

3.4 Existence of a rational choice

We saw earlier in this section that pessimists may have no rational choice from a finite choice set. In this section I show that optimists always do. The simplest case is that of independent acts. By Proposition 5 choosing one such act biases beliefs only over that act. It follows that for an optimistic agent any efficient act must be a rational choice if the acts in the choice set are all independent.

More generally, however, there may be complicated logical dependencies between acts, so that choosing f may actually increase the bias on some other act g more than on f. I therefore take an indirect approach to the general claim. The intuition of the proof is that if

[10] A strictly mixed equilibrium $\sum_i p_i x_i$ can be seen as the mixing of x_i and $\sum_{j \neq i} p_j x_k/(1-p_i)$

the agent is optimistic any strictly mixed equilibrium is unstable (Proposition 7). It follows that the equilibrium with the maximum proportion of some alternative cannot be a mixed one.

Proposition 8. *An optimistic agent has a rational choice from any finite choice set.*

3.5 Comparison with loss-aversion

Kőszegi and Rabin (2006, 2007) develop a model of choice with reference-dependent preferences, in which loss-aversion makes multiple strict choice equilibria possible. The basic phenomenon of multiple equilibria is similar to optimistic choice, but the underlying reason is different, and the circumstances in which it is observed also differ. In particular, loss-aversion never results in the choice of dominated alternatives, but it can be a choice equilibrium for an optimistic agent to choose a strictly dominated option, such as an investment associated with a first-order stochastically dominated payoff distribution. On the other hand, (subjective) uncertainty is necessary for optimistic multiple equilibria, but not for multiple equilibria arising from loss-aversion.

A further difference is that Kőszegi and Rabin (2006, 2007) interpret utility as comparable between different reference points, and define a notion of equilibrium which maximises expected utility across all reference-points. In this paper utility is simply a representation of preferences, and a comparable notion cannot be defined.

4 Non-normative belief updating

Beliefs in the model depend not only on information but also on payoff. Thus they may change not only through the acquisition of relevant information but also from a change in payoffs. Section 3 focused on one way this can happen: choosing an act determines the payoffs in each state and so different choices can lead to different beliefs. In this section, I assume that the payoffs in each state are fixed, and consider instead the effect of new information. The key idea is that fixing the mapping between payoffs and *states* nonetheless leaves room for information to change the relationship between payoffs and *events*. News therefore can affect beliefs not only directly by providing relevant evidence, but also *indirectly*, by mediating the effects of belief distortion.

I start by considering two examples. I then proceed to a formal analysis, proving a proposition that characterises situations in which information leads to a predictable change in beliefs about a second set of events without providing any relevant information about the likelihood of those events. I conclude the section by relating this analysis to psychological evidence of cognitive dissonance and other motivated beliefs change.

Example 2. Picnic

Consider again the picnic example. In Section 3 I looked at how beliefs over "rain" and "shine" depend on the decision whether to go out on a picnic. Here I consider a variation of this example in which the payoff in these events still depends on the activity (picnic or home), but that which activity takes place depends not on choice, but on an exogenous event such as whether or not a friend recovers from an illness. If the friend recovers the picnic is on, and if the friend remains ill the picnic is off. In this situation payoff is defined over combinations of rain/shine and ill/well. The agent knows what payoff she would obtain in each of these four possible combinations, but it is only when uncertainty about the health of the friend is resolved that the agent knows what payoff she would obtain in the events of rain or shine. I assume the same payoff matrix as in Section 3 on the assumption that the picnic is on if and only if the friend recovers:

	Rain	Shine
Well	0	1
Unwell	0.5	0.5

As in the discussion of this example in Section 3, Equation 3 implies that beliefs over the probability of rain are distorted if the picnic is on but not if the agent stays at home. Since the choice of activity is up to the unrelated outside event, it follows that the beliefs the agent ends up with about the weather depend on normatively irrelevant news. Moreover, the agent's beliefs are consistent, and so beliefs prior to learning the news about the friend's health lie between the two possible posteriors. The subjective probability of rain therefore moves with the news about the outside event. If the friend recovers, it goes down. If the friend remains ill, it goes up.

To see further how this works, consider the following numeric example. Suppose each of the four possible combinations of ill/well and rain/shine have an undistorted probability of 1/4, and suppose the agent is optimistic with $e^{\psi} = 4$. Belief distortion is then as follows:[11]

	Rain	Shine	
Well	1/4	1/4	1/2
Unwell	1/4	1/4	1/2
	1/2	1/2	

$\Rightarrow$

	Rain	Shine	
Well	1/9	4/9	5/9
Unwell	2/9	2/9	4/9
	1/3	2/3	

[11] For example, the distorted probability of well+shine is given by $e^{2\psi}/(e^{0\psi} + e^{2\psi} + 2e^{\psi}) = 4/9$.

Note first that while the two sets of events are independent given the undistorted beliefs, they are *not* independent given the distorted beliefs. For example, the subjective probability of rain depends on whether or not the friend recovers. Second, beliefs conditional on friend recovering are biased, but beliefs conditional on the friend staying ill are not. This follows from the fact that the payoff in the event of picnic varies as a function of the weather, but the payoff in the event of staying at home is constant. Third, because of the bias, subjective expected utility conditional on the risky event (picnic) is higher than subjective expected utility conditional on the safe event (staying at home). This parallels the result from Section 5.1 on the preference of optimists for risky choices. Here, however, it is not up to the agent to choose between these two alternatives. Instead, the higher expected utility of the risky alternative leads to a bias in the probability that the picnic would be on: the distorted probability of picnic is 5/9, as compared with an undistorted probability of 1/2. Fourth, prior beliefs over the event of rain are related to posterior beliefs by the law of iterated expectations, just as in the case of standard Bayesian agents[12]. Finally, following the news that the friend recovers, the subjective probability of rain *drops* from 1/3 to 1/5. If, instead, the news is that the friend remains ill, the subjective probability of rain *rises* to 1/2. These changes in belief occur despite the fact that the news about the friend's health carries no real information about the weather.

The key to the picnic example is that good weather is more important if the friend recovers (and hence the picnic is on; in the example it is *only* important if the picnic is on). In other words, good weather and a picnic are *complements*. Consider a second example, this time with substitutes:

Example 3. Sales target

A salesperson is one sale short of a target leading to promotion with two prospects left. The two possible sales are therefore substitutes, and for simplicity I assume they are perfect substitutes:

	B fails	B works
A fails	0	1
A works	1	1

I assume independent undistorted prior beliefs of 1/4 for each of the two possible sales to work, and for illustration, extreme optimism. On this assumption the likelihood of neither

[12]
$$\pi(\text{rain}) = \pi(\text{rain}|\text{picnic}) \cdot \pi(\text{picnic}) + \pi(\text{rain}|\text{home}) \cdot \pi(\text{home}) = \frac{1}{5} \cdot \frac{5}{9} + \frac{1}{2} \cdot \frac{4}{9} = \frac{1}{3} \tag{14}$$

sale working is zero, and that of one or both sales working is 1. Undistorted and distorted beliefs are therefore as follows:

	B fails	B works	
A fails	9/16	3/16	3/4
A works	3/16	1/16	1/4
	3/4	1/4	

$\Rightarrow$

	B fails	B works	
A fails	0	3/7	3/7
A works	3/7	1/7	4/7
	3/7	4/7	

Thus, the undistorted probability of the first sale working is 1/4, but the distorted probability is a much higher 4/7. However, if the salesperson learns that the first sale is a success, then she no longer cares about the prospects of the second, and so the subjective probability of the second deal working drops back to 1/4.

4.1 Formal analysis

The claim I seek to prove is that beliefs about some variable X are shifted in a predictable direction by good news about another variable Y that is a complement or substitute of X. The general setting I consider assumes that utility can be written as a separable function of three random variables, so that $u(s) = u(X, Y, Z) = u(X, Y) + u(Z)$, where X and Y are complements or substitutes. For example, the utility function of the sales agent is given by $u(X, Y) = \max(X, Y)$ for binary $X, Y \in \{0, 1\}$, and X and Y are substitutes. A more general version of this example in which $u = \max(X_1, X_2, \ldots, X_n)$ can be put in the above framework by defining $X = X_1$ and $Y = \max(X_2, \ldots, X_n)$, so that $u(X_1, \ldots, X_n) = \max(X, Y)$. Finally, consider an example such as $u = \sum_i a_i X_i$. Then a_i and X_i are strict complements, so that good news about a_i should raise beliefs about X_i if the agent is optimistic. This example can be put into the framework by defining $X = X_i$, $Y = a_i$ and $Z = \sum_{j \neq i} a_j X_j$.

I denote the undistorted probability distribution of a random variable X by $P_X(x)$ and its distorted probability distribution by $\Pi_X(x)$. I denote the corresponding probability distributions conditional on information I by $P_{X|I}(\cdot)$ and $\Pi_{X|I}(\cdot)$. The proposition is about how news that is about one variable can nonetheless affect beliefs about another unrelated variable. I define news about a variable as information that is independent of other variables, and "good news" as news that changes beliefs about the variable it is about in a predictable direction.

Definition 7 (News about a variable)**.** Suppose X is a random variables then an event I is *news about* X if for all random variables Y that are independent of X relative to $p(\cdot)$ and for all $y \in \mathbb{R}$ the event $Y = y$ is independent of I relative to $p(\cdot)$. If, furthermore, $P_{X|I} \succeq_{LR} P_X$ then I is *good news about* X. If the inequality is strict then I is *strictly* good news about X.

I define complements and substitutes by adapting a standard definition based on increasing (decreasing) differences (Edgeworth, 1925; Samuelson, 1974; Topkis, 1998) that avoids any assumptions on differentiability:

Definition 8 (Complements and substitutes)**.** Suppose $u = u(X, Y) + u(Z)$ as above. Then X and Y are *complements* (*substitutes*) if for all $t > t'$ in the image of X_2, $u(x, t) - u(x, t')$ is non-decreasing (non-increasing) as a function of x. X and Y are *strict complements (substitutes)* if the differences $u(x, t) - u(x, t')$ are strictly increasing (strictly decreasing) as a function of x.

The main claim now follows. I define the proposition for the case of optimism and complements, but the result generalises for substitutes and pessimism with a sign inversion for both changes. The effect on X of good news about Y is summarised in the following table:

	Optimist	Pessimist
Complements	+	-
Substitutes	-	+

Proposition 9. *Suppose $\psi > 0$, X and Y are complements, and I is good news about Y. Then $\Pi(X|I) \succeq_{LR} \Pi(X)$. Moreover, if (i) X and Y are strict complements, (ii) I is strictly good news about Y, and (iii) the probability distribution of X and Y is not degenerate, then the result is also strict.*

The intuition for this result can be obtained by considering the case of two binary variables. The claim is then that $\pi(X = 1|I)/\pi(X = 0|I) \geq \pi(X = 1)/\pi(X = 0)$. Now,

$$\begin{aligned}\frac{\pi(X=1|I)}{\pi(X=0|I)} &= \frac{\pi(X=1, Y=1|I) + \pi(X=1, Y=0|I)}{\pi(X=0, Y=1|I) + \pi(X=0, Y=0|I)} \\ &= \frac{p(X=1, Y=1|I)e^{\psi u(1,1)} + p(X=1, Y=0|I)e^{\psi u(1,0)}}{p(X=0, Y=1|I)e^{\psi u(0,1)} + p(X=0, Y=0|I)e^{\psi u(0,0)}} \\ &= \frac{p(X=1)}{p(X=0)} \cdot \frac{p(Y=1|I)e^{\psi u(1,1)} + p(Y=0|I)e^{\psi u(1,0)}}{p(Y=1|I)e^{\psi u(0,1)} + p(Y=0|I)e^{\psi u(0,0)}}\end{aligned} \tag{15}$$

with a similar result for $\pi(X = 1)/\pi(X = 0)$. Taking the difference between the two several terms drop out, and what remains is

$$\Big(p(Y=1|I)p(Y=0) - p(Y=1)p(Y=0|I)\Big)\Big(e^{\psi(u(1,1)+u(0,0))} - e^{\psi(u(0,1)+u(1,0))}\Big) \tag{16}$$

The expression on the left is weakly positive as I is good news about Y, and the expression on the right is weakly positive as X and Y are complements and $\psi > 0$[13].

4.2 Motivated cognition and cognitive dissonance

The prediction of this section is that information that leads to a change in payoffs can cause belief change even when it is not normatively relevant. This corresponds to the essence of the phenomenon known as *cognitive dissonance* (Festinger and Carlsmith, 1959; Cooper and Fazio, 1984), and sometimes known as *motivated cognition* (Kunda, 1990). One way to summarise the key finding of this literature is that the expected value of a variable is increasing in the payoff for high values of the variable in question. This finding fits the predictions of this section on the assumption that most people are optimistic.

For example, Klein and Kunda (1992) found that subjective beliefs about the likelihood playing with or against a given player in a trivia contest influences beliefs about the ability of that player. Letting x denote the ability of the player, and $a \in \{-1, 1\}$ the side on which she will play, the agent's payoff can be modelled by $u(a, x) = ax$. Thus, a and x are *complements*. The prediction would therefore be that an increase in a (telling subjects that the player will be on their team) should lead to a higher assessment of x (an increased subjective valuation of the ability of the player). This prediction agrees with the empirical findings.

Similarly, Berscheid et al. (1976) found that expecting to date a person causes an increased valuation of that person. Here we can let x_i denote the attractiveness of person i, and p_i the probability that person i is to be the date. Then p_i and x_i are complements, and news that $p_i = 1$ is predicted to increase the bias over x_i, again in agreement with empirical findings.

5 Choice applications

In this section I consider cases where a standard agent sees one risky act f' as better than another risky act f, but where a biased agent may nonetheless view f as better. In terms of choice behaviour suppose g is a safe (constant) act, and that given undistorted beliefs $f \succeq g$ implies $f' \succeq g$. The question I ask is under what conditions is it also the case that $f \succeq_f g$ implies $f' \succeq_{f'} g$, and under what conditions can it be that $f \succeq_f g$, but $f' \prec_{f'} g$?

I start by showing that optimists may prefer a "worse" act in expected utility terms if its payoff is more uncertain. Optimists are therefore *risk-loving*. Similarly I show that pessimists

[13] $e^{\psi(u(1,1)+u(0,0))} - e^{\psi(u(0,1)+u(1,0))} \geq 0$ if and only if $e^{\psi(u(1,1)+u(0,0))} \geq e^{\psi(u(0,1)+u(1,0))}$, and since ln is a monotonic function, this is true if and only if $\psi(u(1,1) + u(0,0) - u(1,0) - u(0,1)) \geq 0$.

are *risk-averse*. I then show that first order stochastic dominance can also be violated. In an application increased punishment can actually *reduce* deterrence. Finally, I show that, unlike the case with first-order stochastic dominance, stochastic dominance in the likelihood ratio *is* a sufficient condition to ensure that $f \succeq_f g$ implies that also $f' \succeq_{f'} g$.

5.1 Uncertainty and risk aversion

In this section I focus on acts with normally distributed payoffs. Suppose X_f has an undistorted normal distribution with mean μ and variance σ^2, and that $g = c$ for some $c \in \mathbb{R}$. Then by Proposition 2 $f \succeq_f g \iff \mu + \psi\sigma^2 \geq c$. In particular, for an optimistic agent $f \succeq_f g \iff \sigma^2 \geq (c - \mu)/\psi$, and for a pessimistic agent $f \succeq_f g \iff \sigma^2 \leq (\mu - c)/|\psi|$.

Thus, if the agent is optimistic and f is uncertain then (i) it can be that $f \succeq_f g$ even if $\mathcal{E}(f) < \mathcal{E}(g)$, and (ii) if the agent is more uncertain about f than about f' then it is possible that $f \succeq_f g$ but $f' \prec_{f'} g$ even if $\mathcal{E}(f') > \mathcal{E}(f)$. An optimistic agent is thus *risk-loving*. Similarly, if the agent is pessimistic and f is uncertain then (i) it may be that $f \prec_f g$ even if $\mathcal{E}(g) < \mathcal{E}(f)$, and (ii) if there is more uncertainty in f' than in f then it is possible that $f \succeq_f g$ but $f' \prec_{f'} g$ even if $\mathcal{E}(f') > \mathcal{E}(f)$. A pessimistic agent is thus *risk-averse*.

This relationship between optimism and risk-loving on the one hand, and pessimism and risk-aversion on the other, suggests the possibility that agents take risks not because they are *tolerant* of the actual risks they are facing, but because they *underestimate* them. Similarly, pessimists who take a cautious approach may do so not because of having a particularly curved utility function, but because they overestimate the probability of losses. These two sources of observed risk preferences can be separated by controlling for the *type* of uncertainty the agent is facing. Optimism and pessimism are only relevant for subjective uncertainty, and have no effect if uncertainty is objective.

5.2 Violation of first-order stochastic dominance

In this section I use an economics of crime example to show that payoff-dependent belief distortion can lead to a violation of first-order stochastic dominance. In the example, an increase in the punishment of criminals can potentially *increase* the expected utility of crime and thereby reduce deterrence. The intuition is that the probability of staying out of jail and level of punishment are *complements*. Therefore, by the results of Section 4 an increase in punishment reduces the subjective probability of getting caught. There are therefore two effects working opposite directions (Figure 3) and with the right parameters an increase in punishment can lead to a counter-intuitive result.

Example 4. Crime and punishment
An optimistic criminal can choose a safe act g of payoff $= -b$ (working in McDonald's) or a risky life of crime f^c, which yields 0 if the criminal stays out of jail and $-c$ in the event, A, that the criminal is caught. c denotes the level of punishment. The claim is that an increase in c (lowering the true expected utility of crime) can nonetheless increase the subjective expected utility of crime, and thereby reduce deterrence.

With these definitions F^c first-order stochastically dominates $F^{c'}$ if $c < c'$. A standard agent who chooses crime if $f^{c'}$ is offered always chooses crime if f^c is offered. Using Equation 3 the perceived odds ratio for A is $\pi_{f^c}(A)/\pi_{f^c}(A^c) = p(A)e^{-\psi c}/p(A^c)$, and the subjective expected utility of crime is given by

$$\mathcal{E}_{f^c}(f^c) = -c\pi_{f^c}(A) = -\frac{p(A)e^{-\psi c}}{p(A)e^{-\psi c} + (1 - p(A))} \cdot c \tag{17}$$

Obviously $\mathcal{E}_{f^c}(f^c)$ is negative for all $c > 0$, and since $\lim_{c\to\infty} e^{-\psi c}c = 0$ for $\psi > 0$ it follows that $\lim_{c\to\infty} \mathcal{E}_{f^c}(f^c) = 0$. Thus, there is a certain critical point c^*, such that increasing punishment beyond c^* actually increases the subjective expected utility of crime, and therefore *reduces deterrence*. As $b < 0$ it follows that for c large enough $f^c \succeq_{f^c} g$. In other words, for c high enough it is a choice equilibrium for the agent to choose crime, even if there are lower values of c for which crime is an irrational choice.

Going back to Equation 17 note that an increase in $p(E)$ unambiguously makes $\mathcal{E}_{f^c} f(c)$ more negative, and so improves deterrence. There is thus an asymmetry: an increase in the probability of arrest is predicted to be more effective than an increase in the severity of punishment, and the latter can even be counter-productive. The evidence on this topic is interesting. For example, Grogger (1991) looking at the frequency of arrests found a much larger deterrent effect for the certainty of punishment as compared with severity (-0.562 with t-score of 8.52 for the probability of conviction vs. 0.017 with t-score of 1.65 for average sentence length). Similarly, Nagin and Pogarsky (2001) also found that the certainty of punishment was a far more robust a deterrent than severity. Note that the effects of belief distortion may be particularly strong if optimists self-select a life of crime over less risky alternatives.

5.3 Stochastic dominance in the likelihood ratio

Though belief distortion can lead to violations of first-order stochastic dominance, the following proposition shows that it respects stochastic dominance in the likelihood ratio:

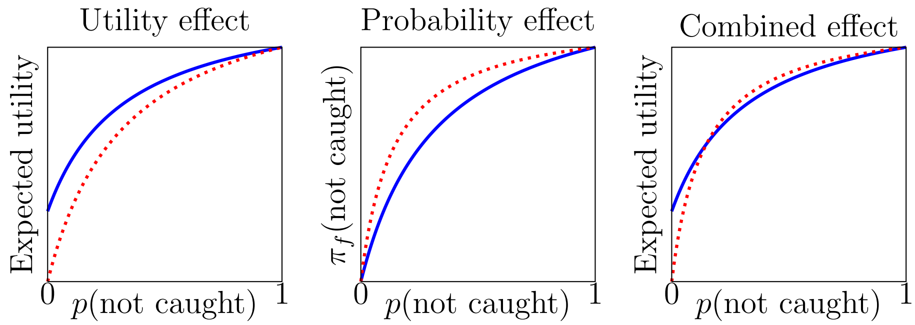

Figure 3: An example of a violation of first-order stochastic dominance: increased severity of punishment *reducing* deterrence. The first panel shows the drop in expected utility that would follow if the subjective probability of arrest remained the same. The second panel shows the increase in the subjective probability of staying out of jail that follows from the increased stakes that result from the increase in punishment. Finally, the third panel shows the combined effect: the plot for a more severe punishment (dotted line) is lower if the probability of arrest is high, but is arrest is sufficiently rare, the increase in punishment actually increases subjective expected utility.

Proposition 10. *Let f and f' be two acts, such that $F' \succeq_{LR} F$. Then for all constant acts g, $f \succeq_f g \Rightarrow f' \succeq_{f'} g$.*

The intuition for this result comes from considering the case of binary bets which differ only in the probability of the desirable outcome. The distorted probabilities are monotonic in this probability, and so the ranking of such bets is invariant to the distortion. Stochastic dominance in the likelihood ratio can be seen as a generalisation of this concept, as the odds ratio between *any* two payoffs is higher in the stochastically dominant act.

6 Conclusion

This paper introduces a model of Bayesian decision making where a person's beliefs about the likelihood of different outcomes depend upon the anticipated payoff consequences of those outcomes. This dependence is modelled as a time-invariant distortion linking the actual beliefs of a person to the beliefs the person would have held if beliefs depended only on information. The fact that the distortion is time-invariant implies that beliefs at time t depend only on time t information and payoffs, and are independent of the order in which payoff was determined or made clear, and the order in which information arrives. In particular, the same beliefs result if (1) payoffs are determined first, and then information arrives, in which case the belief distortion may be seen as an optimistic/pessimistic bias in interpreting information, (2) information arrives first, and then a decision is made, in which case the belief distortion makes choice into an equilibrium phenomenon, and (3) information arrives, but payoffs remain uncertain. Then information about the *payoffs* arrives. This last case results in non-normative belief updating following the arrival of information about the payoffs.

The model accounts for the key evidence for both optimism and pessimism. It is consistent with the comparative statics for the strength of available evidence on the one hand, and the stakes in the outcome on the other. The predictions for possible multiple choice equilibria for optimists and no rational choices for pessimists are supported by evidence for extreme optimism and pessimism. The predictions of non-normative belief updating are supported by cognitive dissonance evidence (Cooper and Fazio, 1984), as well as evidence from studies of "motivated cognition" Kunda (1990).

The model is structured as an extension of the standard Anscombe-Aumann choice model (Anscombe and Aumann, 1963), and can be readily adapted for use in applications. Optimistic bias has been applied to a variety of economic areas, such as financial markets, corporate finance, bargaining, and insurance. One would hope that the model of this paper can lead to better and stronger predictions in some areas. In financial markets in particular there are phenomena that look strongly suggestive of the types of mechanisms explored in this paper. Most obviously, optimistic investors may discount risks. In addition, the dynamic predictions of the model may help explain such phenomena as inefficient lack of diversification, and traders avoiding the sale of poorly performing securities.

The model predicts that an early bet on some event leads to belief distortion over the event. In particular, if the agent is optimistic, then a decision to bet on A increases the subjective likelihood of A. Suppose the agent then receives more information, and has the opportunity to make a second bet, then the early bet on A has the effect of increasing the likelihood that the agent bets on A again. This effect bears on the first-impressions effect, in that the early information that lead into the first bet on A ends up having more impact on the agent's beliefs than later information. The same effect also bears on the sunk-cost effect, in that the early investment represents a sunk cost by the time the second investment is made, and yet it impacts the second investment decision. On the second front, it seems particularly relevant to finance. In corporate finance the result would be an inefficient escalation of business decisions. In financial markets the result would be an inefficient continual holding of under-performing assets, and an inefficient lack of diversification.

Another obvious area for expanding the model is strategic interaction, where there are existing models in the limit of extreme optimism (Yildiz, 2007), but not for more realistic levels of optimism or pessimism. Strategic interaction raises a number of interesting theoretical issues. For example, in isolated strategic interactions optimistic players may both believe the other is making a mistake (they agree to disagree). This, however, cannot be the case if the interaction is repeated to the point that both agents know the true equilibrium payoff. The most interesting case may be an intermediate one, in which the payoff in each interaction is a combination of a fixed factor (which agents can learn), and a second factor that varies between interactions (and which agents are optimistic about). There are also interesting issues to do with interactions between optimistic and unbiased agents and interactions involving pessimists.

Appendix: Proofs

Lemma 1

Proof. If $p(B) = 0$ the claim follows from absolute continuity. Otherwise, let $r = p(B)/p(A)$, and consider first the case that r is rational. By the non-atomicity there is a rational number q, such that $p(B) = mq$ and $p(A) = nq$ for some natural numbers $m, n \in \mathbb{N}$. By non-atomicity and the assumption that $B \subseteq A$, we can write A as the disjoint union of n events $C_1, \ldots, C_n$, s.t. $p(C_i) = q$ for all i, and B is the disjoint union of $C_1 \ldots C_m$. By absolute continuity, $\pi_f(C_i) > 0$ for all i. Let C_i and C_j denote two of these events, then by applying state-independence to C_i and C_j with $g = f$ we have $\pi_f(C_i) = \pi_f(C_j)$. As a probability measure, π_f is additive, and so $\pi_f(B)/\pi_f(A) = m/n = p(B)/p(A)$.

Suppose now that r is irrational. Then there is an increasing sequence of rational numbers $\{q_n\}_{n=1}^{\infty}$, and a decreasing sequence of rational numbers $\{q'_n\}_{n=1}^{\infty}$, such that $\lim_{n=\infty} q_n = \lim_{n=\infty} q'_n = r$. By non-atomicity, there are events $\{B_n\}_{i=1}^{\infty}$ and $\{B'_n\}_{i=1}^{\infty}$, such that for all n, $B_n \subset B_{n+1} \subset B \subset B'_{n+1} \subset B'_n$, $p(B_n)/p(A) = q_n$, and $p(B'_n)/p(A) = q'_n$. By the claim for the rational case it follows that $\pi_f(B_n)/\pi_f(A) = q_n$ and $\pi_f(B'_n)/\pi_f(A) = q'_n$. Taking the limit of $n \to \infty$ we thus obtain that $\pi_f(B)/\pi_f(A) = r$ as required. □

Lemma 2

Proof. Let $f' = aAbBc$, $g' = bAaBc$, and $h' = bAcBa$, and let $D = A \cup B \cup C$. By state independence $\pi_{g'}(A|D) = \pi_{f'}(B|D)$ and $\pi_{g'}(B|D) = \pi_{f'}(A|D)$, and since $C = D \setminus (A \cup B)$ it also follows that $\pi_{g'}(C|D) = \pi_{f'}(C|D)$. Similarly, $\pi_{h'}(A|D) = \pi_{g'}(A|D)$, $\pi_{h'}(B|D) = \pi_{g'}(C|D)$, and $\pi_{h'}(C|D) = \pi_{g'}(B|D)$. Combining the two results, we obtain $\pi_{h'}(A|D) =$

$\pi_{f'}(B|D)$, and $\pi_{h'}(B|D) = \pi_{f'}(C|D)$. Thus,

$$\frac{\pi_{f'}(C|D)}{\pi_{f'}(B|D)} = \frac{\pi_{h'}(B|D)}{\pi_{h'}(A|D)} = \frac{\pi'_h(B|A \cup B)}{\pi_{h'}(A|A \cup B)} = \frac{\pi'_f(B|A \cup B)}{\pi_{f'}(A|A \cup B)} = \frac{\pi'_f(B|D)}{\pi_{f'}(A|D)} \tag{18}$$

where the second step follows from information-independence, the third from shift invariance, and the fourth from information-independence. Finally, by information-independence $\pi_f(\cdot|D) = \pi_{f'}(\cdot|D)$. Combining this with Equation 18 we obtain

$$\frac{\pi_f(C)}{\pi_f(B)} = \frac{\pi_{f'}(C|D)}{\pi_{f'}(B|D)} = \frac{\pi_{f'}(B|D)}{\pi_{f'}(A|D)} = \frac{\pi_f(B)}{\pi_f(A)} \tag{19}$$

□

Lemma 3

Proof. Suppose first that $b - a$, $b' - a'$, and $a' - a$ are all rational numbers. There is then a rational number $q \in \mathbb{Q}$, and natural numbers $k, l, m, n \in \mathbb{N}$, such that $b = a + kq$, $b' = a' + lq$, $a' = a + mq$ and $\max(b, b') = a + nq$. Let $r = \min(p(A), p(B), p(A'), p(B'))/(n + 1)$. By non-atomicity A can be expressed as the disjoint union of events $E_0, E_2, \ldots, E_n$, such that $p(E_i) = r$ for all i. Furthermore, there are events $B_r \subseteq B, A'_r \subseteq A'$ and $B'_r \subseteq B'$ such that $p(B_r) = p(A'_r) = p(B'_r) = r$. Define $F_k = B_r, F_m = A'_r$, $F_{l+m} = B'_r$ and $F_i = E_i$ for $\in \{0, 1, \ldots n\} \setminus \{k, m, l + m\}$. With this construction $p(F_i) = r$ for all i, and $F_0 \subseteq A$, $F_k \subseteq A_k$, $F_m \subseteq A'$ and $F_{l+m} \subseteq B'$ even if A, A', B and B' are not all distinct.

Let g denote the act defined by $g(F_i) = a + iq$ for $0 \leq i \leq n$ then $g(F_{i+1}) - g(F_i) = q$ for all i, and so by Lemma 2, there is a constant $t \in \mathbb{R}$, s.t. $\pi_g(F_{i+1})/\pi_g(F_i) = t$ for all i. For any $j \in \{0, 1, \ldots, n\}$, $\pi_g(F_j)/\pi_g(F_0) = (\pi_g(F_j)/\pi_g(F_{j-1})) \cdots (\pi_g(F_1)/\pi_g(F_0)) = t^j$.

Let $D = F_0 \cup F_k \cup F_m \cup F_{l+m}$. By construction $F_0 = A \cap D, F_k = B \cap D, F_m = A' \cap D$, and $F_{l+m} = B' \cap D$. Moreover, $g(s) = f(s)$ for all $s \in D$, so that by information independence $\pi_f(\cdot|D) = \pi_g(\cdot|D)$. Thus, $\pi_f(B \cap D)/\pi_f(A \cap D) = \pi_f(B|D)/\pi_f(A|D) = \pi_g(B|D)/\pi_g(A|D) = \pi_g(F_k)/\pi_g(F_0) = t^k$, and similarly $\pi_f(B' \cap D)/\pi_f(A' \cap D) = t^l$.

Finally, by Lemma 1, $\pi_f(A \cap D) = (p(A \cap D)/p(A))\pi_f(A) = (r/p(A))\pi_f(A)$. Using the result from the previous paragraph we obtain $\pi_f(B)/\pi_f(A) = (p(B)/p(A))(\pi_f(B \cap D)/\pi_f(A \cap D)) = (p(B)/p(A))t^k$, and similarly, $\pi_f(B')/\pi_f(A') = (p(B')/p(A'))t^l$. Thus,

$$\frac{\pi_f(B')}{\pi_f(A')} \cdot \frac{p(A')}{p(B')} = t^l = (t^k)^{ql/qk} = \left(\frac{\pi_f(B)}{\pi_f(A)} \cdot \frac{p(A)}{p(B)}\right)^{\frac{b'-a'}{b-a}} \tag{20}$$

and the identity in Equation 5 is obtained by taking logs. Finally, the general case in which $b - a$, $b' - a'$, and $a' - a$ are not necessarily all rational follows from prize continuity. □

Proposition 1

Proof. That any logit distortion is consistent is obvious. For the other direction, note first that part 1 of the definition of logit distortion is simply absolute continuity. For the second part, consider first the case where f is a constant act, or that otherwise there is a prize $a \in I_f$, such that $Supp(p) \subseteq f^{-1}(a)$. In this case, for any events A and B in the definition of logit distortion it must be the case that $b = a$, and so by Lemma 1,

$$\ln \frac{\pi_f(B)}{\pi_f(A)} = \ln \frac{\pi_f(B)/\pi_f(f^{-1(a)})}{\pi_f(A)/\pi_f(f^{-1(a)})} = \ln \frac{p(B)/p(f^{-1}(a))}{p(A)/p(f^{-1}(a))} = \ln \frac{p(B)}{p(A)} + \psi(b - a) \qquad (21)$$

for all $\psi \in \mathbb{R}$. Suppose now that there are at least two different prizes $a, b \in I_f$, such that $p(f^{-1}(a)) > 0$ and $p(f^{-1}(b) > 0$. Let $a^* < b^*$ denote any two such prizes, define $A^* = f^{-1}(a^*)$ and $B^* = f^{-1}(b^*)$, and set $\psi = \ln(\pi_f(B^*)/\pi_f(A^*)) - \ln(p(B^*)/p(A^*))$.

Let now A and B denote any two events, such that $p(A) > 0$, $p(B) > 0$, $A \subseteq f^{-1}(a)$, and $B \subseteq f^{-1}(b)$ for some $a, b \in I_f$. Suppose first that $b > a \geq a^*$. Then by Lemma 3,

$$\begin{aligned} \ln \frac{\pi_f(B)}{\pi_f(A)} - \ln \frac{p(B)}{p(A)} &= \ln \left(\frac{\pi_f(B)}{\pi_f(A)} \cdot \frac{p(A)}{p(B)} \right) = \frac{b-a}{b^* - a^*} \cdot \ln \left(\frac{\pi_f(B^*)}{\pi_f(A^*)} \cdot \frac{p(A^*)}{p(B^*)} \right) \\ &= \frac{b-a}{b^* - a^*} \cdot (\psi(b^* - a^*)) = \psi(b - a) \end{aligned} \qquad (22)$$

The other cases ($a < a^*$ and/or $b < a$) are very similar. □

Proposition 2

Proof. Let $f_f(x) = dF_f(x)/dx$ denote the probability density function of the payoff of f under the distorted probability measure π_f. Then,

$$\begin{aligned} f_f(x) &\propto \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) e^{\psi x} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2 - 2\psi\sigma^2 x}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\{\mu+\psi\sigma^2\})^2 - 2\psi\sigma^2\mu - \psi^2\sigma^4}{2\sigma^2}} \propto \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\{\mu+\psi\sigma^2\})^2}{2\sigma^2}} \end{aligned} \qquad (23)$$

$F_f(x)$ is thus a normal distribution with mean $\mu + \sigma^2\psi$ and variance σ^2. □

Proposition 3

Proof. By Equation 1 $dF_f^{\psi}(x)/dF_f^{\psi'}(x) = (p(x)e^{\psi x})/(p(x)e^{\psi' x}) = e^{(\psi-\psi')x}$ which is a non-decreasing function of x if $\psi \geq \psi'$. The claim thus follows from the definition of stochastic dominance in the likelihood ratio. □

Proposition 4

Proof. The proof is immediate from Equation 3 by considering the events A_{min} and A_{min}^c and taking the limit $\psi \to -\infty$, and similarly for the events A_{max} and A_{max}^c with the limit $\psi \to \infty$. □

Proposition 5

Proof. The proof follows from Equation 3 and the assumption that X_f and X_g are independent relative to the undistorted probability measure p:

$$G_f(x) = \pi_f(X_g \leq x) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} \pi_f(X_f = u, X_g = v)dudv = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{x} p(X_f = u, X_g = v)e^{\psi u}dudv}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(X_f = u, X_g = v)e^{\psi u}dudv}$$
$$= \frac{\left(\int_{-\infty}^{\infty} p(X_f = u)e^{\psi u}du\right)\left(\int_{-\infty}^{x} p(X_g = v)dv\right)}{\left(\int_{-\infty}^{\infty} p(X_f = u)e^{\psi u}du\right)\left(\int_{-\infty}^{\infty} p(X_g = v)dv\right)} = G(x) \tag{24}$$

□

Proposition 6

Proof. I prove the first part of the claim, as the second part is very similar. Let f and g be as in the claim, let $a^* = \max_{s\in Supp(p)} g(s)$, and let $A^* = g^{-1}(a^*)$ denote the event that a^* is obtained. By Proposition 4, $\lim_{\psi\to\infty} \pi_g(A^*) = 1$. Thus, $\lim \psi \to \infty \mathcal{E}_g(g) = a^*$. But by proposition 5 $\mathcal{E}_g(f) = \mathcal{E}(f) < \max_{s\in Supp(p)} g(s) = a^*$. Thus, $\lim_{\psi\to\infty} \mathcal{E}_g(g) - \mathcal{E}_g(f) \geq 0$, and so for ψ large enough $g \succeq_g f$. □

Proposition 7

Proof. Note that the preference $g \succeq_{f(\epsilon)} h$ equals the sign of the difference in expected utilities, which is a differentiable function of ϵ. It follows that an equilibrium is stable (unstable) if and only if the derivative of this utility difference is non-positive (non-negative) when equated at

$\epsilon = 0$. Now, if $g(s) = h(s)$ for all $s \in Supp(p)$ then both g and h are rational. Otherwise,

$$\begin{aligned}
\text{sign}\left(\frac{\partial}{\partial \epsilon}\mathcal{E}_f(g(s) - h(s))\right)\Big|_{\epsilon=0} &= \text{sign}(\frac{\partial}{\partial \alpha'}\frac{\sum_{s\in S} p(s)(g(s)-h(s))e^{\psi((\alpha' g(s)+(1-\alpha')h(s))}}{\sum_{s\in S} p(s)e^{\psi(\alpha' g(s)+(1-\alpha')h(s))}})\Big|_{\alpha'=\alpha} \\
&= \text{sign}(\frac{\partial}{\partial \alpha'}\sum_{s\in S} p(s)(g(s)-h(s))e^{\psi(\alpha' g(s)+(1-\alpha')h(s))})\Big|_{\alpha'=\alpha} \\
&= \text{sign}(\psi \sum_{s\in S} p(s)(g(s)-h(s))^2 e^{\psi(\alpha' g(s)+(1-\alpha')h(s))})\Big|_{\alpha'=\alpha} \\
&= \text{sign}(\psi)
\end{aligned} \tag{25}$$

where the second inequality follows from $g \sim_f h$ for a strictly mixed equilibrium f[14]. □

Proposition 8

Proof. The proof is by induction on the size, n, of the finite choice set C. The claim is obviously true for $n = 1$. For $n = 2$ let $C = \{g, h\}$ and define $f(\alpha) = \alpha g + (1-\alpha)h$, and $\rho(\alpha) = \sum_{s\in S} \pi_{f(\alpha)}(g(s) - h(s))$, the expected utility difference between g and h with $f(\alpha)$ as the reference. $\rho(\alpha)$ is a continuous function of α. If $\rho(1) \geq 0$ then $g \in C$ is a rational choice, and if $\rho(0) \leq 0$ then $h \in C$ is a rational choice. Suppose, contrary to the claim, that neither is the case. By the intermediate value theorem there is then $\alpha^* \in (0,1)$ such that $\rho(\alpha^*) = 0$ and $\rho(\alpha) < 0$ for all $\alpha \in (\alpha^*, 1]$. Now, $f(\alpha^*)$ is a fully mixed rational choice in $C' = \{f(\alpha)\}_{\alpha\in[0,1]}$, and by Proposition 7 is an unstable equilibrium. Thus, there is $\alpha \in [\alpha^*, 1]$ such that $\rho(\alpha) > 0$, a contradiction to the definition of α^*.

Finally, in the general case there is (by the standard fixed point argument) at least one rational act $f = \sum_{i=1}^n p_i x_i$ where $x_i \in C$ and $\sum_{i=1}^n p_i = 1$. Without loss of generality suppose $p_1 > 0$. If $p_1 = 1$ then $f \in C$, and the agent has a rational choice in C. Otherwise, let $g = \sum_{i=2}^n (p_i/(1-p_1))x_i$. With this definition $f = p_1 x_1 + (1-p_1)g$ and so x_1 is a rational choice by the proof for the $n = 2$ case. □

Proposition 10

Proof. I prove the stronger claim that $F_f \succeq_{LR} F_{f'}$. Let b and b be two possible prizes with $b > a$. I need to prove that $\pi_f(X_f = b)/\pi_f(X_f = a) \geq \pi_{f'}(X_{f'} = b)/\pi_{f'}(X_{f'} = a)$. By Equation 3 this is true if and only if $(p(X_f = b)/p(X_f = a))e^{\psi(b-a)} \geq (p(X_{f'} = b)/p(X_{f'} =$

[14] This is a standard result.

$a))e^{\psi(b-a)}$, i.e. if $p(X_f = b)/p(X_f = a) \geq p(X_{f'} = b)/p(X_{f'} = a)$. But this last inequality follows from the assumption that $F \succeq_{LR} F'$. □

Proposition 9

Proof. I need to prove that for all pairs of possible values $x > x'$ for X, $\pi(X = x|I)/\pi(X = x'|I) \geq \pi(X = x)/\pi(X = x')$. Now,

$$\begin{aligned}
\frac{\pi(X=x|I)}{\pi(X=x'|I)} &= \frac{\sum_y \sum_z \pi(X=x, Y=y, Z=z|I)}{\sum_y \sum_z \pi(X=x', Y=y, Z=z|I)} \\
&= \frac{\sum_y \sum_z p(X=x, Y=y, Z=z|I) e^{\psi u(x,y,z)}}{\sum_y \sum_z p(X=x', Y=y, Z=z|I) e^{\psi u(x',y,z)}} \\
&= \frac{p(X=x)}{p(X=x')} \cdot \frac{\sum_y \sum_z p(Y=y) p(Z=z) e^{\psi(u(x,y)+u(z))}}{\sum_y \sum_z p(Y=y) p(Z=z) e^{\psi(u(x',y)+u(z))}} \\
&= \frac{p(X=x)}{p(X=x')} \cdot \frac{\sum_y p(Y=y|I) e^{\psi u(x,y)}}{\sum_y p(Y=y|I) e^{\psi u(x',y)}}
\end{aligned} \tag{26}$$

where the second step uses Equation 6, the third step the assumption that X, Y and Z are independent, that I is news about Y, and that $u(x, y, z) = u(x, y) + u(z)$. Similarly,

$$\frac{\pi(X=x)}{\pi(X=x')} = \frac{p(X=x)}{p(X=x')} \cdot \frac{\sum_{y'} p(Y=y') e^{\psi u(x,y')}}{\sum_{y'} p(Y=y') e^{\psi u(x',y')}} \tag{27}$$

Thus the claim is true if and only if

$$\frac{\sum_y p(Y=y|I) e^{\psi u(x,y)}}{\sum_y p(Y=y|I) e^{\psi u(x',y)}} - \frac{\sum_{y'} p(Y=y') e^{\psi u(x,y')}}{\sum_{y'} p(Y=y') e^{\psi u(x',y')}} \geq 0 \tag{28}$$

or

$$\sum_y \sum_{y'} p(Y=y|I) p(Y=y') \cdot \left(e^{\psi(u(x,y)+u(x',y'))} - e^{\psi(u(x,y')+u(x',y))}\right) \geq 0 \tag{29}$$

This expression is antisymmetric in y and y'. The terms with $y = y'$ therefore drop out, and terms with $y' > y$ can be combined with the corresponding term for which $y' < y$. Thus, the following condition is equivalent:

$$\begin{aligned}
&\sum_y \sum_{y'<y} \big(p(Y=y|I) p(Y=y') - p(Y=y'|I) p(Y=y)\big) \\
&\qquad \cdot \left(e^{\psi(u(x,y)+u(x',y'))} - e^{\psi(u(x,y')+u(x',y))}\right) \geq 0
\end{aligned} \tag{30}$$

Now, $p(Y = y|I)p(Y = y') - p(Y = y'|I)p(Y = y)$ is non-negative by the fact that $y' < y$ and the assumption that I is good news about y, and $e^{\psi(u(x,y')+u(x',y))}$ is non-negative by the assumption that $\psi \geq 0$, X and Y are complements, and the fact that e^x is a monotonically increasing function. It therefore follows that the entire expression is non-negative, thereby concluding the main proof.

Finally, note that if $\psi > 0$, I is strictly good news about Y and X and Y are strict complements, then all the expressions are strictly positive. The condition that X and Y are non-degenerate ensures that the claim is not empty, and that the sums contain at least one term. When all these conditions hold the inequality is therefore strict. □

References

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

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*, American Psychiatric Publishing.

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

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

Beck, A. (1967). *Depression: Clinical, Experimental, and Theoretical Aspects*, Imprint unknown.

Berscheid, E., Graziano, W., Monson, T. and Dermer, M. (1976). Outcome dependency: Attention, attribution, and attraction, *Journal of Personality and Social Psychology* **34**(5): 978–988.

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

Cooper, J. and Fazio, R. (1984). A new look at dissonance theory, *Advances in experimental social psychology* **17**: 229–266.

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

Edgeworth, F. (1925). The Pure Theory of Monopoly, *Papers Relating to Political Economy* **1**: 111–142.

Festinger, L. and Carlsmith, J. (1959). Cognitive consequences of forced compliance., *J Abnorm Psychol* **58**(2): 203–10.

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.

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

Kőszegi, B. and Rabin, M. (2007). Reference-dependent risk attitudes, *American Economic Review* **97**(4): 1047–1073.

Klein, W. and Kunda, Z. (1992). Motivated person perception: constructing justifications for desired beliefs, *Journal of experimental social psychology* **28**(2): 145–168.

Kunda, Z. (1990). The case for motivated reasoning, *Psychological Bulletin* **108**(3): 480–498.

Luce, D. (1959). *Individual Choice Behavior*, Wesley, New York.

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

Samuelson, P. (1974). Complementarity, *Journal of Economic Literature* **12**: 1255–1289.

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

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

Spiegler, R. and Eliaz, K. (2008). Consumer Optimism and Price Discrimination, *Theoretical Economics (forthcoming)* .

Topkis, D. (1998). *Supermodularity and Complementarity*, Princeton University Press.

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

Yildiz, M. (2004). Waiting to persuade, *The Quarterly Journal of Economics* **119**(1): 223–248.

Yildiz, M. (2007). Wishful Thinking in Strategic Environments, *Review of Economic Studies* **74**(1): 319–344.