

Mental Processes and Decision Making*

Olivier Compte
PSE, Paris

Andrew Postlewaite
University of Pennsylvania

March 2009; this version February 2011

*Compte: Paris School of Economics, 48 Bd Jourdan, 75014 Paris (e-mail: compte@enpc.fr); Postlewaite: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6207 (e-mail: apostlew@econ.upenn.edu). A large part of this work was done while Postlewaite visited the Paris School of Economics; their hospitality is gratefully acknowledged. We thank Eddie Dekel, David Dillenberger, Jeff Ely, Mehmet Ekmekci, Kfir Eliaz, Yuval Salant and Alvaro Sandroni for helpful discussions, as well as the participants of numerous seminars at which the paper was presented. The authors thank the Gould Foundation for financial support, and Postlewaite thanks the National Science Foundation for financial support.

Abstract

Consider an agent who is unsure of the state of the world and faces computational bounds on mental processing. The agent receives a sequence of signals imperfectly correlated with the true state that he will use to take a single decision. The agent is assumed to have a finite number of “mental states” that quantify his beliefs about the relative likelihood of the states, and uses the signals he receives to move from one state to another. At a random stopping time, the agent will be called upon to make a decision based solely on his mental state at that time. We show that under quite general conditions it is optimal that the agent ignore signals that are not very informative, that is, signals for which the likelihood of the states is nearly equal. We suggest this model as an explanation of systematic inference mistakes people sometimes make.

It ain't so much the things we don't know that get us into trouble.
It's the things that we know that just ain't so.
Artemus Ward

1 Introduction

Robert is convinced that he has ESP and offers the following statement to support this belief: “I was thinking of my mother last week and she called right after that.” Robert is not alone in his beliefs; more people believe in ESP than in evolution, and in the U.S. there are twenty times as many astrologers as there are astronomers.¹ Readers who don't believe in ESP might dismiss Robert and other believers as under-educated anomalies, but there are sufficiently many other similar examples to give pause. Nurses who work on maternity wards believe (incorrectly) that more babies are born when the moon is full², and it is widely believed that infertile couples who adopt a child are subsequently more likely to conceive than similar couples who did not adopt (again, incorrectly).³

We might simply decide that people who hold such beliefs are stupid or gullible, at the risk of finding ourselves so described for some of our own beliefs.⁴ Whether or not we are so inclined, many positive economic models have at their core a decision-making module, and those models must somehow take account of agents' actual beliefs, however unsound we may think them.

Our interest in the widespread belief in ESP goes beyond the instrumental concern for constructing accurate decision making modules for our models. The deeper question is why people hold such questionable beliefs? The simple (simplistic?) response that a large number of people are stupid is difficult to accept given the powerful intellectual tools that evolution has provided us in many domains. How is that evolution has generated a brain that can scan the symbols on a page of paper and determine which subway connected to which bus will systematically get someone to work on time, and yet believe in ESP?

Our aim in this paper is to reconcile the systematic mistakes we observe in the inferences people draw from their experiences with evolutionary forces that systematically reward good decisions. We will lay out a model of how an individual processes streams of informative signals that is (a) optimal, and (b) leads to incorrect beliefs such as Robert's. The reconciliation is possible because of computational bounds we place on mental processing. Roughly speaking, our restrictions on mental processing preclude an agent from recalling every signal he receives perfectly. Consequently, he must employ some sort of summary

¹See Gilovich (1991), page 2.

²See G. O. Abell and B. Greenspan (1979).

³See E. J. Lamb and S. Leurgans (1979).

⁴There are numerous examples of similarly biased beliefs people hold. Research has demonstrated that people frequently estimate the connection between two events such as cloud seeding and rain mainly by the number of positive-confirming events, that is, where cloud seeding is followed by rain. Cases of cloud seeding and no rain and rain without cloud seeding tend to be ignored (Jenkins and Ward (1965) and Ward and Jenkins (1965).)

statistics that capture as well as possible the information content of all the signals that he has seen. We assume that agents do not have distinct mental processes for every problem they might face, hence an optimal process will do well for “typical” problems, but less well for “unusual” problems.⁵ Given the restrictions agents face in our model, they optimally ignore signals that are very uninformative. Robert’s experience of his mother calling right after he thought of her is quite strong evidence in support of his theory that he has ESP. His problem lies in his having not taken into account the number of times his mother called when he *hadn’t* thought of her. Such an event may have moved Robert’s posterior belief that he had ESP only slightly, but the accumulation of such small adjustments would likely have overwhelmed the small number of instances which seem important. Our primary point is that the mental processing property that we suggest leads Robert to conclude that he has ESP – ignoring signals that by themselves have little information – will, in fact, be optimal when “designing” a mental process that must be applied to large sets of problems when there are computational bounds.

We lay out our model of mental processes in the next section. The basic model is essentially that analyzed by Wilson (2004) and by Cover and Hellman (1970), but our interest differs from that of those authors. In those papers, as in ours, it is assumed that an agent has bounded memory, captured by a set of mental states. Agents receive a sequence of signals that are informative about which of two possible states of Nature is the true state. The question posed in those papers is how the agent can optimally use the signals to move among the finite set of mental states, knowing that at some random time he will be called upon to make a decision, and his current mental state is all the information he can use about the signals he has received.

Cover and Hellman and Wilson characterize the optimal way to transit among mental states as additional signals arrive when the expected number of signals the agent receives before making a decision goes to infinity. Our interest differs from these authors in two respects. First, as mentioned above, our point of view is that an agent’s mental system – the set of mental states and the transition function – have evolved to be optimal for a class of problems rather than being designed for a single specific problem. Second, and more importantly, we are interested in the case that the expected number of signals that an agent will receive before making a decision is *not* necessarily large.

We compare different mental systems in section 3 and discuss the implications of our main theorems in section 4. We discuss our analysis in section 5.

Related literature

The central theme of this paper is that a decision maker uses a single decision

⁵Early papers that investigated a decision maker who uses a single decision protocol for a number of similar but not identical problems include Baumol and Quandt (1964). A more systematic modelling of the idea can be found in Rosenthal (1993), where it is assumed that an agent will choose among costly rules of thumb that he will employ in the set of games they face. Lipman (1995) provides a very nice review of the literature on modelling bounds on rationality resulting from limits on agents’ ability to process information.

protocol for a number of similar but not identical problems. There is some literature addressing this issue going back at least to Baumol and Quandt (1964). A more systematic modelling of the idea can be found in Rosenthal (1993), where it is assumed that an agent will choose among costly rules of thumb that he will employ in the set of games they face. Lipman (1995) provides a very nice review of the literature on modelling bounds on rationality resulting from limits on agents' ability to process information. These papers

2 The model

Decision problem. There are two states, $\theta = 1, 2$. The true state is $\theta = 1$ with probability π^0 . An agent receives a sequence of signals imperfectly correlated with the true state, that he will use to take a single decision. The decision is a choice between two alternatives, $a \in \{1, 2\}$. To fix ideas, we assume the following payoff matrix, where $g(a, \theta)$ is the payoff to the agent when he takes action a in state θ :

$$\begin{array}{cc} g(a, \theta) & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \end{array}$$

There are costs c_1 and c_2 associated with decisions 1 and 2 respectively. Let $c = (c_1, c_2)$ denote the cost profile, and $u(a, \theta, c)$ the utility associated with each decision a when the state is θ and cost profile is c . We assume that the utility function takes the form

$$u(a, \theta, c) = g(a, \theta) - c_a.$$

The cost c is assumed to be drawn from a distribution with full support on $[0, 1] \times [0, 1]$. The cost vector c is known to the agent prior to the decision. It is optimal to choose $a = 1$ when the agent's belief $\pi \geq \frac{1+c_1-c_2}{2}$. In what follows, we let $v(\pi)$ denote the payoff the agent derives from optimal decision making when π is his belief that the true state is $\theta = 1$. We have:

$$v(\pi) = E_{c_1, c_2} \max\{\pi - c_1, 1 - \pi - c_2\}.$$

It is straightforward to show that v is strictly convex. For example, if costs are uniformly distributed, $v(\pi) = v(1 - \pi)$ and a calculation shows that for $\pi \geq 1/2$, $v(\pi) = \frac{1}{6} + \frac{4}{3}(\pi - \frac{1}{2})^2(2 - \pi)$ (see figure 1).

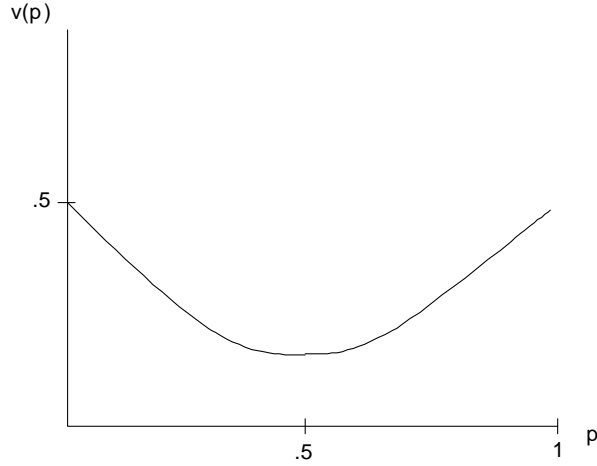


Figure 1: $v(\pi)$ when costs are distributed uniformly

The signals received. Signals are drawn independently, conditional on the true state θ , from the same distribution with density $f(\cdot | \theta)$, assumed to be positive and smooth on its support. When signal x arises, there is a state $\theta(x)$ that has highest likelihood, namely:

$$\theta(x) = \arg \max_{\theta} f(x | \theta)$$

It will be convenient to denote by $l(x)$ the likelihood ratio defined by:

$$l(x) = \frac{f(x | \theta = \theta(x))}{f(x | \theta \neq \theta(x))}.$$

The state $\theta(x)$ is the state for which signal x provides support, and the likelihood ratio $l(x)$ provides a measure of the strength of the evidence in favor of $\theta(x)$. We assume that the set of signals x that are not informative (i.e. $l(x) = 1$) has measure 0.

We assume that signals are received over time, at dates $t = 0, 1, \dots$, and that the decision must be taken at some random date $\tau \geq 1$. For simplicity, we assume that τ follows an exponential distribution with parameter $1 - \lambda$:

$$P(\tau = t | \tau \geq t) = 1 - \lambda.$$

This assumption captures the idea that the agent will have received a random number of signals prior to making his decision. The fact that this number is drawn from an exponential distribution is not important, but makes computations tractable. The parameter λ provides a measure of the number of signals the agent is likely to receive before he must make a decision: the closer λ is to 1, the larger the expected number of signals. Note that the agent always receives at least one signal.

We assume that that players correctly interpret the signals they see. Our result that it is optimal for agents to ignore weakly informative signals is robust to agents making perception errors in which they sometimes incorrectly perceive the strength of the evidence they see, and sometimes incorrectly perceive which theory the evidence supports.⁶

Limited information processing. A central element of our analysis is that agents cannot finely record and process information. Agents are assumed to have a limited number of *states of mind*, and each signal the agent receives is assumed to (possibly) trigger a change in his state of mind. We have in mind, however, that those transitions apply across many decision problems that the agent may face, so the transition will not be overly problem-specific or tailored to the particular decision problem at hand. Thus, we shall assume that transitions may only depend on the perceived likelihood ratio associated with the signal received. Formally a state of mind is denoted $s \in S$, where S is a finite set. For any signal x received, changes in state of mind depend on the perceived likelihood ratio \tilde{l} associated with x . We denote by T the transition function:

$$s' = T(s, \tilde{l}).$$

To fix ideas, we provide a simple example. We will later generalize the approach.

Example 1: The agent may be in one of three states of mind $\{s_0, s_1, s_{-1}\}$. His initial state is s_0 . When he receives a signal x , he gets a perception $(\tilde{\theta}, \tilde{l})$. The event $A_0^+ = \{\tilde{\theta} = 1\}$ corresponds to evidence in favor of state $\theta = 1$, while $A_0^- = \{\tilde{\theta} \neq 1\}$ corresponds to evidence against $\theta = 1$. Transitions are as follows:

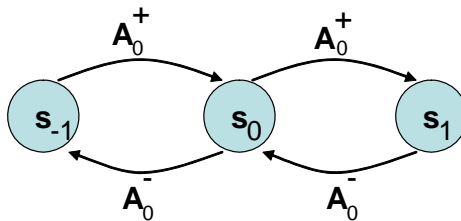


Figure 2: Transition function

As the above figure illustrates, if the agent finds himself in state s_1 when he is called upon to make his decision, there may be many histories that have led to his being in state s_1 . We assume that the agent is limited in that he is unable to distinguish more finely between histories. Consequently, S and T are simply devices that generate a particular pooling of the histories that the agent faces when making a decision.

Evidence

⁶See Compte and Postlewaite (2010b) for a discussion.

Optimal behavior. Our aim is to understand the consequences of limits on mental states and exogenous transitions among those states. To focus on those aspects of mental processing, we assume that the agent behaves optimally contingent on the state he is in. We do not claim that there are not additional biases in the way agents process the information they receive; indeed, there is substantial work investigating whether, and how, agents may systematically manipulate the information available to them.⁷

Formally, the set of mental states S , the initial state, the transition function T and Nature's choice of the true state generate a probability distribution over the mental states the agent will be in in any period t that he might be called upon to make a decision. These distributions along with the probability distribution over the periods that he must make a decision, determine a probability distribution over the state the agent will be in when he makes a decision, $q^{S,T}(s)$. This along with the probability distribution over the true state θ determines a joint distribution over (s, θ) , denoted $q^{S,T}(s, \theta)$.

We assume that the agent is able to identify the optimal decision rule $a(s, c)$ ⁸; that is, that the agent can maximize:

$$\sum_{s, \theta} q^{S,T}(s, \theta) E_c u(a(s, c), \theta, c).$$

Call $\pi(s) = \Pr\{\theta = 1 \mid s\}$ the Bayesian updated belief. For convenience, we will sometimes use the notation $\pi_\theta(s) = \Pr\{\theta \mid s\}$. Since $q^{S,T}(s, \theta) = \pi_\theta(s)q^{S,T}(s)$, the expected utility above can be rewritten as:

$$\sum_s q^{S,T}(s) \sum_\theta \pi_\theta(s) E_c u(a(s, c), \theta, c).$$

This expression is maximal when the agent chooses $a = 1$ when $\pi_1(s) - c_1 \geq \pi_2(s) - c_2$, and the maximum expected utility can be rewritten as

$$V(S, T) = \sum_s q^{S,T}(s) v(\pi(s)).$$

We illustrate our approach next with specific examples.

Computations. We will illustrate how one computes the distribution over states prior to decision making. Define

$$p_\theta = \Pr\{\tilde{\theta} = 1 \mid \theta\}.$$

⁷Papers in this area in psychology include Festinger (1957), Josephs et al. (1992) and Sedikides et al. (2004). Papers in economics that pursue this theme include Benabou and Tirole (2002, 2004), Brunnermeier and Parker (2005), Compte and Postlewaite (2004), and Hvide (2002).

⁸It is indeed a strong assumption that the agent can identify the optimal decision rule. As stated above, our aim is to demonstrate that even with the heroic assumption that the agent can do this, he will systematically make mistakes in some problems.

Denote by ϕ the distribution over states of mind at the time the agent makes a decision⁹:

$$\phi = \begin{pmatrix} \phi(s_1) \\ \phi(s_0) \\ \phi(s_{-1}) \end{pmatrix}.$$

Also denote by ϕ^0 the initial distribution over states of mind (i.e., that puts all weight on s_0 , so that $\phi^0(s_0) = 1$); Conditional on the true state being θ , one additional signal moves the distribution over states of mind from ϕ to $M_\theta\phi$, where

$$M^\theta = \begin{pmatrix} p_\theta & p_\theta & 0 \\ 1 - p_\theta & 0 & p_\theta \\ 0 & 1 - p_\theta & 1 - p_\theta \end{pmatrix},$$

is the transition matrix associated with the mental process (S, T) .

Starting from ϕ^0 , then conditional on the true state being θ , the distribution over states of mind at the time the agent takes a decision will be:

$$\phi_\theta = (1 - \lambda) \sum_{n \geq 0} \lambda^n (M^\theta)^{n+1} \phi^0$$

or equivalently,

$$\phi_\theta = (1 - \lambda)(I - \lambda M^\theta)^{-1} M^\theta \phi^0. \quad (1)$$

These expressions can then be used to compute $q^{S,T}(s, \theta)$, $q^{S,T}(s)$ and $\pi_\theta(s)$. For example, $q^{S,T}(s, 1) = \pi^0 \phi_1(s)$ and $q^{S,T}(s) = \pi^0 \phi_1(s) + (1 - \pi^0) \phi_2(s)$.

More generally, given any mental process (S, T) , one can associate a transition matrix M^θ that summarizes how an additional signal changes the distribution over states of mind when the true state is θ , and then use (1) to derive the distribution over states $q^{S,T}(s)$ and the Bayesian posteriors $\pi_\theta(s)$, and subsequently, expected welfare $v(S, T)$.

The fully symmetric case.

We will illustrate our ideas with specific symmetric density functions given θ . Suppose that $x \in [0, 1]$ and $f(x | \theta = 1) = 2x$.¹⁰ Figure 3 shows the density functions of signals for each of the two states $\theta = 1, 2$.

⁹ ϕ is the conditional distribution on S given $\theta = 1$, that is, $\phi(s) = q^{S,T}(s | \theta = 1)$.

¹⁰Given our symmetry assumption, this implies $f(x | \theta = 2) = 2(1 - x)$. So signals x above $1/2$ are evidence in favor of $\theta = 1$, and the strength of the evidence ($l(x) = \frac{x}{1-x}$) gets large when x gets close to 1.

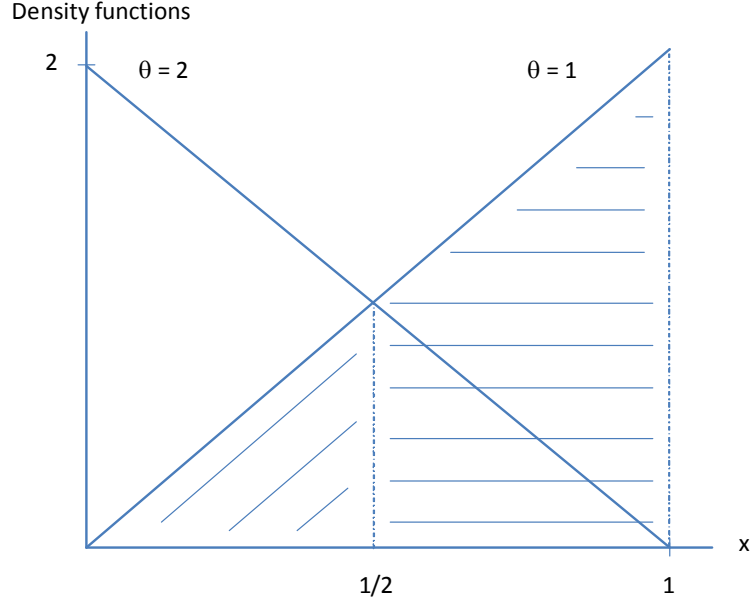


Figure 3: State contingent density functions

A signal $x < 1/2$ is evidence in favor of $\theta = 2$ while a signal $x > 1/2$ is evidence in favor of $\theta = 1$. If $\theta = 1$ is the true state the horizontally shaded region below the density function given $\theta = 1$ and to the right of $1/2$ represents the probability of a signal in favor of $\theta = 1$, while the diagonally shaded region to the left of $1/2$ represents the probability that the signal is “misleading”, that is, of a signal in favor of $\theta = 2$. The probability of “correct” and “wrong” signals are respectively $3/4$ and $1/4$. Thus, if the decision maker is in the mental state that is associated with $\theta = 1$, s_1 , there is a $1/4$ chance that the next signal will entail his leaving that mental state. If the expected number of signals is very large (that is, λ is close to 1), the probability distribution over his mental states will be close to the ergodic distribution on the states s_{-1}, s_0, s_1 when $\theta = 1$ is $(1/13, 3/13, 9/13)$. Consequently, the probability that the decision maker’s mental state indicates the correct θ is a bit less than $3/4$.

We turn now to a general symmetric signal structure and assume that $\pi_0 = 1/2$.

Assumption 1: $x \in [0, 1]$, $f(x | \theta = 1) = f(1 - x | \theta = 2)$.

Under assumption 1, $\Pr\{\tilde{\theta} = 1 | \theta = 1\} = \Pr\{\tilde{\theta} = 2 | \theta = 2\}$. We let $p = \Pr\{\tilde{\theta} = 1 | \theta = 1\}$. Note that we must have $p > 1/2$. Given these symmetry assumptions, $\pi(s_0) = 1/2$ and $\pi(s_1) = 1 - \pi(s_{-1})$.

The following proposition gives $\pi(s_1)$ and $q^{S,T}(s_1)$ as a function of λ and p .

Proposition 1: $\pi(s_1) = p \frac{1 - \lambda(1-p)}{1 - 2\lambda p(1-p)}$ and $q^{S,T}(s_1) = \frac{1 - 2\lambda(1-p)p}{2 - 2\lambda^2(1-p)p}$.

As p or λ get closer to 1, beliefs become more accurate (conditional on s_1 or s_{-1}), and there is a greater chance that the agent will end up away from s_0 . At the limit where p is close to 1, the agent almost perfectly learns the correct state before taking a decision. Note that $\pi(s_1) > p$ for all values of λ . Intuitively, being at state of mind $s = s_1$ means that the balance of news in favor/against state s_1 tilts in favor of s_1 by on average more than just one signal. The reason is that if the agent is in state 1, it is because he just received a good signal, and because last period he was either in state 0 (in which case, by symmetry, the balance must be 0) or in state 1 (in which case the balance was already favorable to state s_1).

Finally, note that proposition 1 permits welfare analysis. To get a simple expression, assume that the distribution over costs (c_1, c_2) is symmetric, which implies that $v(\pi) = v(1 - \pi)$. Then expected welfare is:

$$2q^{S,T}(s_1)v(\pi(s_1)) + (1 - 2q^{S,T}(s_1))v\left(\frac{1}{2}\right).$$

As one expects, welfare increases with the precision of the signal (p).

3 Comparing mental processes

Our objective. Our view is that a mental processing system should work well in a variety of situations, and our main interest lies in understanding which mental process (S, T) works reasonably well, or better than others. In this section, we show that there is always a welfare gain to ignoring mildly informative signals.

3.1 An improved mental process

We return to our basic mental system defined in example 1, but we now assume that a signal must be minimally informative to generate a transition, that is, to be taken as evidence for or against a particular state. Formally, we define:

$$A^+ = \{\tilde{\theta} = 1, \tilde{l} > 1 + \beta\} \text{ and } A^- = \{\tilde{\theta} = 2, \tilde{l} > 1 + \beta\}.$$

In other words, the event $A = \{\tilde{l} < 1 + \beta\}$ does not generate any transition. Call (S, T^β) the mental process associated with these transitions. Compared to the previous case ($\beta = 0$), the pooling of histories is modified. We may expect that because only more informative events are considered, beliefs conditional on s_1 or s_{-1} are more accurate. However, since the agent is less likely to experience transitions from one state to another, the agent may have a greater chance of being in state s_0 when making a decision.

We illustrate this basic tradeoff by considering the symmetric case discussed above with $f(x | \theta = 1) = 2x$. Figure 4 below indicates the signals that are ignored for $\beta = 1$: a signal $x \in (1/3, 2/3)$ generate likelihoods in $(1/2, 2)$, and are consequently ignored by the transition T^1 .

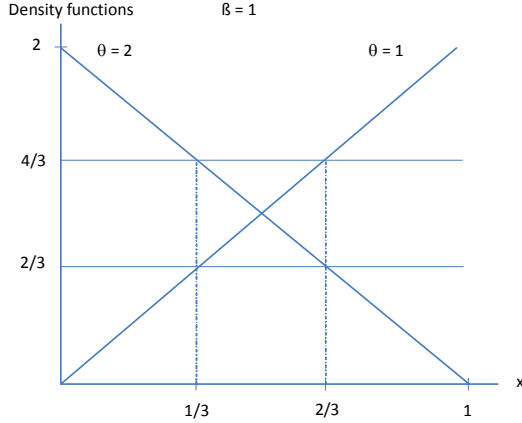


Figure 4

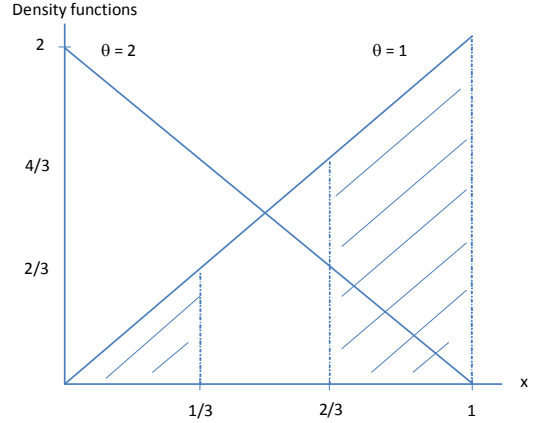


Figure 5

When $\theta = 1$, the shaded regions in Figure 5 to the right of $2/3$ to the left of $1/3$ indicate respectively the probabilities of signals in support of θ_1 and in support of θ_2 , and signals in the interval $(1/3, 2/3)$ are ignored. Now, probabilities of “correct” and “misleading” signals are $5/9$ and $1/9$ respectively, and the ergodic distribution for T^1 is $(1/31, 5/31, 25/31)$. Under T^1 , when λ is close to 1, the probability that the decision maker is in the mental state associated with the true state is nearly $5/6$, as compared with the probability under T^0 , slightly less than $3/4$.

Larger β would lead to even higher probabilities; indeed, as β goes to infinity, the probability of being in the “right” mental state goes to 1. But this increase in accuracy comes at a cost. These probabilities are those associated with the ergodic distribution, but the decision maker will get a finite (random) number signals, and the expected number of signal he will receive before making a decision will be relatively small unless λ is close to 1. When β increases, the probability that the signal will be ignored in any given period goes to 1. Consequently, there is a tradeoff in the choice of β : higher β lead to an increased probability of getting *no* signals before the decision maker must decide, but having more accurate information if he gets *some* signals.

Welfare

Assume that costs are drawn uniformly. We will plot expected welfare as a function of β for various values of λ for the example above.

Each mental process (S, T_β) and state θ generates transitions over states as a function of the signal x . Specifically, let $\alpha = \frac{\beta}{2(2+\beta)}$. When for example the current state is s_0 and x is received, the agent moves to state s_1 if $x > \frac{1}{2} + \alpha$, he moves to state s_2 if $x < \frac{1}{2} - \alpha$, and he remains in s_0 otherwise. Denote by $y = \Pr\{\tilde{\theta} = 1, \tilde{l} < 1 + \beta \mid \theta = 1\}$ and $z = \Pr\{\tilde{\theta} = 2, \tilde{l} < 1 + \beta \mid \theta = 1\}$.¹¹

¹¹ $y = (1/2 + \alpha)^2 - (1/2)^2 = \alpha(1 + \alpha)$, and $z = (1/2)^2 - (1/2 - \alpha)^2 = \alpha(1 - \alpha)$.

Conditional on each state $\theta = 1, 2$, the transition matrices are given by:

$$M_{\beta}^{\theta=1} = \begin{pmatrix} p+z & p-y & 0 \\ 1-p-z & y+z & p-y \\ 0 & 1-p-z & 1-p+y \end{pmatrix}$$

and symmetrically:

$$M_{\beta}^{\theta=2} = \begin{pmatrix} 1-p+y & 1-p-z & 0 \\ p-y & y+z & 1-p-z \\ 0 & p-y & p+z \end{pmatrix}.$$

As before, these matrices can be used to compute beliefs $\pi(s_1)$ and the probability to reach s_1 , that is $q(s_1)$. Increasing β typically raises $\pi(s_1)$ (which is good for welfare), but it makes it more likely to end up in s_0 (which adversely affects welfare). Figure 6 shows how welfare varies as a function of β for two values of λ , $\lambda = 0.5$ (the lower line) and $\lambda = 0.8$ (the upper line). These correspond to expected numbers of signals equal to 2 and 5 respectively.

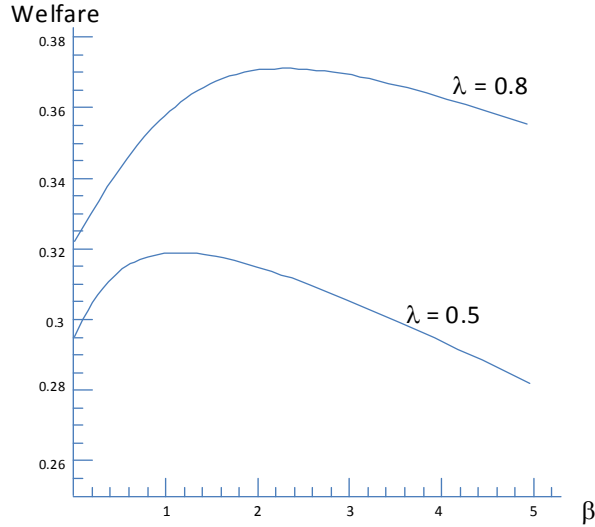


Figure 6: Welfare as a function of β

Note that for a fixed value of λ , for very high values of β there would be little chance of ever transiting to either s_1 or s_2 , hence, with high probability the decision would be taken in state s_0 . This clearly cannot be optimal, so β cannot be too large. Figure 6 also suggests that a value of β set too close to 0 would not be optimal either. The graph illustrates the basic trade-off: large β 's run the risk of never leaving the initial state while small β 's have the agent leaving the "correct" state too easily. When λ is sufficiently large, the first effect is small; consequently, the larger λ , the larger is the optimal value of β .

3.2 Simple mental processes.

The advantage of ignoring weakly informative signals in the example above does not depend on the symmetry assumption, nor on the specific mental system of the example. We first generalize the example to a class of simple mental processes, as defined below.

A simple mental process is described by a set of mental states S and transitions T^0 which specify for each state $s \in S$ and perception $\tilde{\theta} \in \{1, 2\}$ a transition to state $T^0(s, \tilde{\theta})$. Note that the transition depends only on the perception of which state the signal is most supportive of and not on the strength of the signal. We shall restrict attention to mental processes for which T^0 has no absorbing subset. Consider any such simple mental process (S, T^0) . We define a *modified* simple mental process as a simple mental process that ignores weak evidence. Specifically, we define (S, T^β) as the mental process that coincides with (S, T^0) when the perception of the strength of the evidence is sufficiently strong, that is when $\{\tilde{l} > 1 + \beta\}$, and that does not generate a change in the agent's mental state when $\{\tilde{l} < 1 + \beta\}$.

Denote by $W(\lambda, \beta)$ the welfare associated with mental process (S, T^β) , and denote by \underline{W} the welfare that an agent with a single mental state would derive.¹² The next proposition states that for any value of λ , so long as $W(\lambda, 0) > \underline{W}$ and that all states are reached with positive probability, an agent strictly benefits from having a mental process that ignores poorly informative signals.¹³

Proposition 2: Consider a simple mental process (S, T^0) . There exist $a > 0$ and $\beta_0 > 0$ such that for all λ and $\beta \in [0, \beta_0]$:

$$W(\lambda, \beta) - W(\lambda, 0) \geq a\beta\underline{q}(\lambda)[W(\lambda, 0) - \underline{W}],$$

where $\underline{q}(\lambda) = \min_{s \in S} q^{S, T^0}(s)$ denotes the minimum weight on any given state.

Proof: See appendix.

The left hand side of the inequality in the proposition is the welfare increase that results from modifying the transition function (S, T^0) by ignoring signals for which the strength of the evidence is less than β . The proposition states that the gain is a positive proportion of the value of information for the initial transition function.

¹² $\underline{W} = v(\pi_0)$.

¹³In this paper we limit attention to the case that there are two possible theories. Compte and Postlewaite (2010b) discuss the case when there are more than two theories. As one would expect, it continues to be optimal to ignore weakly informative signals in that case. In addition, they argue that agents may tend to see patterns in the data they get when there are no patterns.

3.3 Generalization

A simple mental process does not distinguish mild and strong evidence. A more sophisticated mental process that takes into consideration the strength of the evidence can improve welfare. The following example illustrates this.

Example 2. There are 4 signals, $\{\bar{x}, \bar{y}, \underline{x}, \underline{y}\}$. Signals \bar{x} and \bar{y} are evidence in favor of $\theta = 1$, while signals \underline{x} and \underline{y} are evidence in favor of $\theta = 2$. Signals \bar{x} and \underline{x} are strong evidence, while signals \bar{y} and \underline{y} are mild evidence. Specifically, assume that

$$\tilde{l}(\bar{x}) = \tilde{l}(\underline{x}) = \bar{l} > \tilde{l}(\bar{y}) = \tilde{l}(\underline{y}) = \underline{l} > 1.$$

Finally, we let ν denote the probability of a strong signal.¹⁴ The three numbers $(\nu, \bar{l}, \underline{l})$ fully characterize the distribution over signals.¹⁵

We wish to compare welfare according to whether the agent uses a simple mental process that does not distinguish between mild and strong evidence, and a more sophisticated mental process that would. The two transition functions we consider are as shown in the figures below.

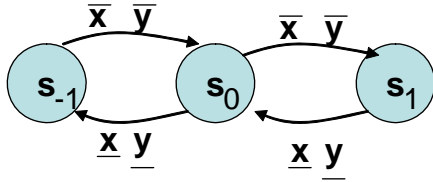


Figure 7: Simple mental process

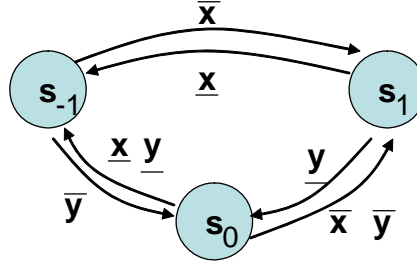


Figure 8: Sophisticated mental process

The sophisticated mental process may lead to higher welfare, but it need not. Intuitively, the sophisticated process should do relatively well when there is a substantial difference between the strong signals (x) and the weak signals (y). When the strengths of the signals are not substantially different, the simple process may outperform the sophisticated process, as we can illustrate with an example. Set $\underline{l} = 2$, $\lambda = 0.9$ and $\nu = 1/10$ (i.e., strong signals are infrequent). Figure 9 below shows welfare as a function \bar{l} for the simple and sophisticated mental processes described above. The sophisticated process gives higher welfare for high levels of \bar{l} , but lower welfare for low levels.

¹⁴This probability is assumed to be independent of the state θ .

¹⁵For example $\Pr\{\bar{x} | 1\} = \frac{\bar{l}\nu}{1+\bar{l}}$, and $\Pr\{\bar{y} | 1\} = \frac{\bar{l}(1-\nu)}{1+\bar{l}}$.

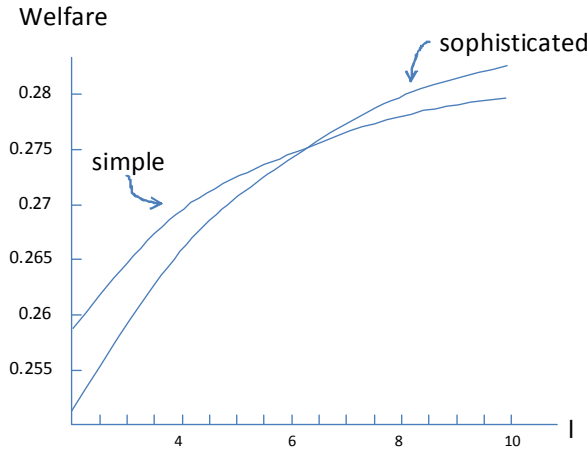


Figure 9: Simple vs. sophisticated transition

It is not only the relative strengths of the evidence that determines which of the two processes leads to higher welfare. Clearly, if the probability of getting a strong signal is sufficiently small, the simple mental process will do better than the sophisticated process, as we show in the next example.

Set $\bar{l} = 8$ and $\underline{l} = 2$ and $\lambda = 0.4$. Figure 10 shows the percent increase in welfare of the sophisticated process over the simple process as a function of the probability of the strong signal, ν . When ν close to 0, the two transitions are equivalent. As ν rises above 0, welfare is higher for the more sophisticated transition. As ν increases above some threshold, the simple transition uses the limited number of states more efficiently, and the simpler transition yields higher welfare.

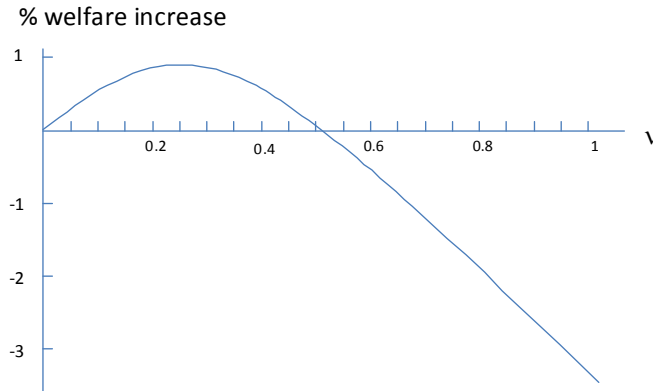


Figure 10: Relative gain with sophisticated transition

These graphs show that there may be an advantage to more sophisticated transitions, though not always. Our next result shows that our main proposition extends to sophisticated mental systems.

We first define a class of *level-k transitions*. Consider an increasing sequence of $(\beta_1, \dots, \beta_{k-1})$, and set $\beta_0 = 0, \beta_k = +\infty$. Any perception $\tilde{l} \in (1 + \beta_{h-1}, 1 + \beta_h)$ with $h \in \{1, \dots, k\}$ is labelled as a perception of strength h . A level-k transition is a transition function for which, given the current state, two perceptions of the same strength generate the same transitions. The agent's perception can thus be summarized by a pair $(\tilde{\theta}, \tilde{k})$. Given any level k mental system for which $T(s, (\tilde{\theta}, \tilde{k})) \neq s$ for some s , one may consider a modified mental system T^β that would coincide with T when $\tilde{l} > 1 + \beta$, but that would ignore signals for which $\tilde{l} < 1 + \beta$.

The following proposition shows the benefit of ignoring weak information. We denote again by $W(\beta, \lambda)$ the welfare associated with the modified mental process. The original level k mental system along with an initial distribution over mental states generates (for each value of λ) posteriors conditional on the state. Denote by $\Delta(\lambda)$ the smallest difference between these posteriors.¹⁶ We have:

Proposition 3: Consider any level k mental system such that $T(s, (\tilde{\theta}, \tilde{k})) \neq s$ for some s . Then, there exists $a > 0$ and $\beta_0 > 0$ such that for all $\beta \in [0, \beta_0]$ and all λ ,

$$W(\beta, \lambda) - W(0, \lambda) \geq a\beta[\Delta(\lambda)]^2 \underline{q}(\lambda),$$

where $\underline{q}(\lambda) = \min_{s \in S} q^{S, T^0}(s)$ denote minimum weight on any given state.

Proof: See Appendix.

4 Consequences of ignoring weak evidence

What are the consequences of the fact that $\beta > 0$ for an agent's transition function? As we discussed in the introduction, our view is that there is a mental process that summarizes the signals an agent has received in a mental state, and that the agent chooses the optimal action given his mental state when he is called upon to make a decision. Agents do not have a different mental process for every possible decision they might some day face. Rather, the mental process that aggregates and summarizes their information is employed for a variety of problems with different signal structures f . β is set optimally *across* a set of problems, not just for a specific distribution over signals f (that in addition would be correctly perceived). When β is set optimally across problems, it means that for some problems, where players frequently receive mild evidence and occasionally strong evidence, there is a bias (in a sense that we make precise below) towards the theory that generates occasional strong evidence, when Bayesian updating might have supported the alternative theory.¹⁷

¹⁶ $\Delta(\lambda) > 0$ except possibly for a finite number of values of λ or initial distribution over states.

¹⁷ Note that we do not have in mind that the decision maker would have biased posterior beliefs. We have assumed throughout the paper that the decision maker is able to maximize

4.1 Casual beliefs

We wish here to be more precise about the exact sense in which a bias towards a particular theory may arise, and to do that we introduce the notion of *casual beliefs*.

Our view is that each mental state reflects, at a casual level, some belief (possibly more or less entrenched) about whether a particular theory is valid. In our basic three mental state example, being in state s_1 is meant to reflect the agent's casual belief that $\theta = 1$ is likely to hold, while s_{-1} is meant to reflect the casual belief that $\theta = -1$ **O: we have to decide whether we want to use θ_{-1} or $\theta = 2$.** is likely to hold; also, s_0 reflects some inability to form an opinion as to which **is the** true state.

One interpretation is that casual beliefs are what the decision maker would report if asked about his inclination as to which state holds. A decision maker would then have *unbiased* casual beliefs if he is more likely to be in mental state s_1 rather than state s_{-1} (hence to lean towards believing $\theta = 1$ rather than $\theta = 2$) whenever the true state is $\theta = 1$. And similarly when the true state is $\theta = 2$.

Not surprisingly, our symmetric example leads to unbiased casual beliefs. In addition, as we saw earlier, increasing β from 0 makes the mental system more accurate: if he does not receive too few messages, the probability that the decision maker is in the mental state associated with the true state increases.

As one moves away from symmetric cases, however, having a positive β may be a source of bias. We discuss one such asymmetric case below.

4.2 An asymmetric example.

We consider an asymmetric case for which most signals are either strongly informative of theory θ_1 , or mildly informative of theory θ_{-1} . We plot below densities that correspond to such a case, with $f(. | \theta_{-1})$ being the uniform distribution on $[0, 1]$.

welfare in choosing an action given his mental state at the time he decides, implying that he behaves as if he had correct posterior beliefs.

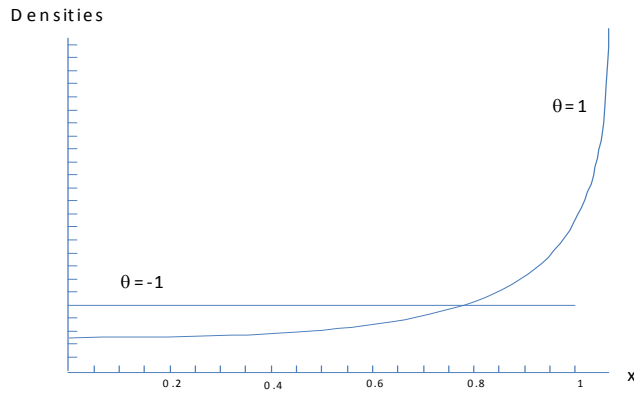


Figure 11: An asymmetric example

Each signal x supports theory $\tilde{\theta} \in \{\theta_1, \theta_{-1}\}$ depending on whether $f(x | \theta_1)$ or $f(x | \theta_{-1})$ is larger (recall that l denotes the strength of the signal). Figure 12 plots the likelihood, l , as a function of x , and, for $\beta = 0.5$, it shows the set signals that support each theory. Signals below .8 are more likely for $\theta = -1$; signals sufficiently low (below approximately .5) generate likelihoods greater than 1.5, hence taken as evidence in support of θ_{-1} . Signals above .8 are more likely for $\theta = 1$, and sufficiently high signals generate likelihoods greater than 1.5 and are taken as evidence in support of θ_1 . The sufficiently informative signals are shown by the thick line in figure 12 below.

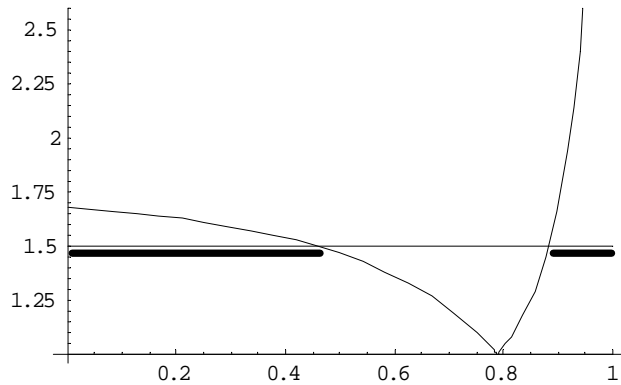


Figure 12: Relevant signals for $\beta = .5$

Suppose now that β is increased to 0.7 (see Figure 13 below). Now, signals are either ignored or come in support of the same theory, θ_1 . Consequently, for an agent whose transition function is characterized by $\beta \geq .7$, only signals in support of θ_1 will be relevant, and should the agents move to s_1 , that mental

state is absorbing.

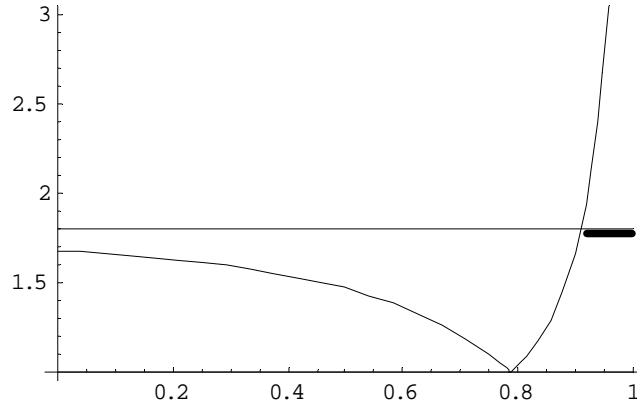


Figure 13: Relevant signals for $\beta = .7$

Although signals may only come in support of theory θ_1 , the agent may not necessarily be in state s_1 when he takes a decision, because he may not have gotten a signal triggering a move to s_1 . In addition, the probability that he gets such a signal depends on the true state, and the probability that he remains in s_0 is thus higher when the true state is θ_{-1} . This illustrates why our basic mental system is beneficial even in this asymmetric setting: posterior beliefs differ at s_0 and s_1 , and welfare is significantly higher than its minimum value, as the following figure illustrates. Figure 14 plots welfare as a function of λ for values of $\beta = 1, 2, 3$ and 4.

Welfare

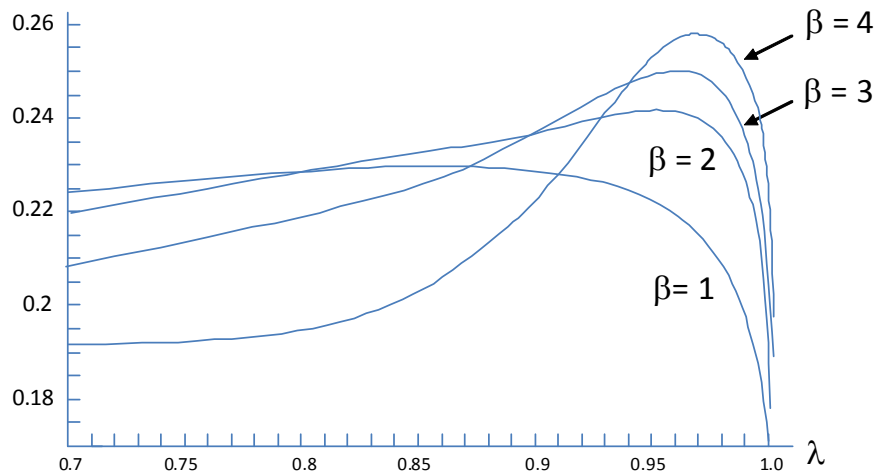


Figure 14: Welfare for different β 's

Note that as λ increases however, and if $\beta > 0.7$, it becomes more likely that the agent will end up in state s_1 whatever the true state, so casual beliefs are biased.

5 Discussion

In our approach, the agent receives signals (x) that interprets support for theories he thinks possible. We formalized this process by defining a transition function that captures how he takes a signal into account. However we do not have in mind that the agent is fully aware of this mental process. The transition function or mental system aggregates the agent's perceptions over time, and the agent's choice is only to choose the optimal behavior conditional for each mental state he may be in when a decision is called for. The transition function is *not* optimized for a specific problem, but rather is assumed to be optimal, on average, for a set of problems that typically differ in the probability distributions over the signals conditional on the true state of nature and in the expected number of signals the agent will receive prior to making a decision. Even with few states, this is already a difficult task. Our interest then is to investigate modifications of the mental system that increase the agent's welfare across over the set of problems he faces.

In comparison, a fully Bayesian agent would have to perform one of the following tasks. If he knew the distribution over signals and errors, he could use these to compute posteriors correctly. Alternatively, he would have to find (through experience?) optimal behavior as a function of all possible histories of signals for each problem that he faces.

Of course, by adding many states and by looking at the transitions (which map states and perceptions to states) which maximize expected welfare for a given problem, one would expect to eventually approximate the decisions that a fully Bayesian agent would make. But given that the number of possible transitions grows exponentially with the number of states, this illustrates how difficult it is to approximate such behavior.

However complex it is to approximate Bayesian behavior, the problem is an order of magnitude more difficult if the transitions must be derived independently for each problem (characterized in our setting by a distribution over signals f and a value of λ) that one faces. Our view is that while deriving optimal behavior at each state could be problem specific, modifications of transitions (i.e., changes in β) should be considered to the extent that they generate welfare increases on average across a set of problems.

6 Appendix

Proof of Proposition 2:

Proof: We fix some initial state, say s_0 . A history h refers to a termination date $\tau \geq 1$ and a sequence of perceptions $((\tilde{\theta}_0, \tilde{l}_0), (\tilde{\theta}_1, \tilde{l}_1), \dots, (\tilde{\theta}_{\tau-1}, \tilde{l}_{\tau-1}))$. Denote by H_s^β the set of histories h that lead to state s (starting from initial state s_0) when the mental process is (S, T^β) . Also denote by $H_{s,s'}$ the set of histories that lead to state s under (S, T^0) and to state s' under (S, T^β) .

For any set of histories H , we denote by $\pi(H)$ the Bayesian posterior, conditional on the event $\{h \in H\}$:

$$\pi(H) = \frac{\Pr(H \mid \theta = 1)}{\Pr(H)} \pi_0.$$

By definition of $W(\lambda, \beta)$, we have:

$$W(\lambda, \beta) = \sum_{s \in S} \Pr(H_s^\beta) v(\pi(H_s^\beta)).$$

Now let \bar{W} denote the welfare that the agent would obtain if he could distinguish between the sets $H_{s,s'}$ for all s, s' . We have:

$$\bar{W} = \sum_{s, s' \in S} \Pr(H_{s,s'}) v(\pi(H_{s,s'})).$$

We will show that there exist constants c and c' , and β_0 such that for all $\beta < \beta_0$,

$$W(\lambda, \beta) \geq \bar{W} - c[\beta \ln \beta]^2 \quad (2)$$

and

$$W(\lambda, 0) \leq \bar{W} - c'\beta. \quad (3)$$

Intuitively, welfare is \bar{W} when the agent can distinguish between all $H_{s,s'}$. Under (S, T^β) , he cannot distinguish between all $H_{s,s'}$:

$$H_s^\beta = \cup_{s'} H_{s',s}.$$

Under (S, T^0) , he cannot distinguish between all $H_{s,s'}$ either, but the partitioning is different:

$$H_s^0 = \cup_{s'} H_{s,s'}. \quad (4)$$

Because each mental process corresponds to a pooling of histories coarser than $H_{s,s'}$, and because v is a convex function, both $W(\lambda, 0)$ and $W(\lambda, \beta)$ are smaller than \bar{W} . What we show below however is that the loss is negligible in latter case (of second order in β), while it is of first order in the former case.

We use three Lemmas, the proofs of which are straightforward. We let $\underline{q} = \min_s \Pr H_s^0$ and $\underline{\Delta} = W(\lambda, 0) - \bar{W}$. We assume that $\underline{q} > 0$. In what follows, we choose β small so that $\gamma(\beta) \equiv \Pr[\tilde{l} < 1 + \beta]$ is small compared to \underline{q} .¹⁸

Lemma 1: There exists a constant c and β_0 such that for all $\beta < \beta_0$ and λ :

$$|\pi(H_{s',s}) - \pi(H_s^\beta)| \leq c\beta |\ln \beta|. \quad (5)$$

Proof: The event $H_{s',s}$ differ from H_s^β only because there are dates t where the perception $(\tilde{\theta}^t, \tilde{l}^t)$ has $\tilde{l}^t < 1 + \beta$. For any fixed $\bar{\lambda} < 1$

¹⁸This is possible because f is smooth and because the event $\{\tilde{l} = 1\}$ has measure 0.

and any $\lambda \leq \bar{\lambda} < 1$, there are a bounded number of such dates, in expectation, so the conclusion follows. For λ close to 1, the number of such dates may become large. However, to determine posteriors up to $O(\beta)$, there are only a number of perceptions prior to the decision comparable to $|\ln \beta|$ that matter.¹⁹ So the conclusion follows as well.²⁰

Lemma 2: There exist constants c' and β_0 such that for all λ and $\beta < \beta_0$, there exist s and $s' \neq s$ with $\Pr(H_{s,s'}) \geq c'\beta$ such that :

$$|\pi(H_{s,s'}) - \pi(H_s^0)| \geq c'[W(\lambda, 0) - \underline{W}]^{1/2}.$$

Proof: We shall say that two states s and s' are consecutive when there is a perception $\tilde{\theta}$ such that $s' = T(s, \tilde{\theta})$. Let $\Delta = W(\lambda, 0) - \underline{W} > 0$. There must exist two states s, s' (not necessarily consecutive) such that $|\pi(H_{s'}^0) - \pi(H_s^0)| \geq c_0\Delta^{1/2}$. Since T has no absorbing subset, there must exist a finite sequence of consecutive states $s^{(0)}, \dots, s^{(k)}, \dots, s^{(K)}$ such that $s^{(0)} = s$ and $s^{(K)} = s'$. Hence there must exist two consecutive states s_0, s'_0 such that $|\pi(H_{s'_0}^0) - \pi(H_{s_0}^0)| \geq c_0\Delta^{1/2}/N$ (where N is the total number of states). The events $H_{s'_0}^0$ and $H_{s_0}^\beta$ both consist of the union of $H_{s'_0, s'_0}$ and of events that have probability comparable to $\gamma(\beta) \equiv \Pr\{\tilde{l} < 1 + \beta\}$. For β small enough, $\gamma(\beta)$ can be made small compared to \underline{q} . The posteriors $\pi(H_{s'_0}^0)$ and $\pi(H_{s_0}^\beta)$ must thus both be close to $\pi(H_{s'_0, s'_0})$ hence close to each other. Applying Lemma 1, it then follows that $|\pi(H_{s_0, s'_0}) - \pi(H_{s_0}^0)| \geq c'\Delta^{1/2}$ for some c' independent of λ . Since s_0 and s'_0 are consecutive, the event H_{s_0, s'_0} must have probability at least comparable to $\gamma(\beta)$, hence at least comparable to β (since f is smooth).

Lemma 3: Let m, \bar{m} such that $\bar{m} \geq v'' \geq m$. For α small, we have, forgetting second order terms in α :

$$\alpha \frac{\bar{m}}{2} (\pi^1 - \pi^0)^2 \geq (1-\alpha)v(\pi^0) + \alpha v(\pi^1) - v(\alpha\pi^0 + (1-\alpha)\pi^1) \geq \alpha \frac{m}{2} (\pi^1 - \pi^0)^2.$$

Since

$$\sum_{s \in S} \Pr(H_{s,s'}) \pi(H_{s,s'}) = \pi\left(\bigcup_s H_{s,s'}\right) \sum_{s \in S} \Pr(H_{s,s'}) = \pi(H_{s'}^\beta) \Pr(H_{s'}^\beta),$$

¹⁹More precisely, consider the last date prior to the decision where the state is s_0 . Because there is a finite number of states and no absorbing subsets, the probability of not going through s_0 during T periods is at most equal to μ^T for some $\mu < 1$. So for T comparable to $|\ln \beta|$, there is a probability $o(\beta)$ to stay away from s_0 . So with probability $1 - o(\beta)$, fewer than $O(|\ln \beta|)$ perceptions matter.

²⁰Note that the bound can be improved, because the expected fraction of the time where $(\tilde{\theta}^t, \tilde{l}^t)$ has $\tilde{l}^t < 1 + \beta$ gets close to 0 with β .

we have:

$$\begin{aligned}\bar{W} - W(\lambda, \beta) &= \sum_{s' \in S} \left[\sum_{s \in S} \Pr(H_{s,s'}) [v(\pi(H_{s,s'})) - v(\pi(H_{s'}^\beta))] \right] \\ &= \sum_{s' \in S} \Pr(H_{s'}^\beta) \left[\sum_{s \in S} \frac{\Pr(H_{s,s'})}{\Pr(H_{s'}^\beta)} [v(\pi(H_{s,s'})) - v(\sum_{s \in S} \frac{\Pr(H_{s,s'})}{\Pr(H_{s'}^\beta)} \pi(H_{s,s'}))] \right].\end{aligned}$$

Applying Lemma 3 thus yields:

$$\bar{W} - W(\lambda, \beta) \leq c \max_{s, \hat{s}, s'} |\pi(H_{s,s'}) - \pi(H_{\hat{s},s'})|^2,$$

and Lemma 1 gives a lower bound on $W(\lambda, \beta)$.

To get the upper bound on $W(\lambda, 0)$, we use:

$$\sum_{s' \in S} \Pr(H_{s,s'}) \pi(H_{s,s'}) = \pi\left(\bigcup_{s'} H_{s,s'}\right) \sum_{s' \in S} \Pr(H_{s,s'}) = \pi(H_s^0) \Pr(H_s^0),$$

and write:

$$\bar{W} - W(\lambda, 0) = \sum_{s \in S} \left[\sum_{s' \in S} \Pr(H_{s,s'}) [v(\pi(H_{s,s'})) - v(\pi(H_s^0))] \right].$$

Lemmas 2 and 3 (using s and s' as defined in Lemma 2) then yield a lower bound:

$$\bar{W} - W(\lambda, 0) \geq c\beta[\pi(H_{s,s'}) - \pi(H_s^0)]^2.$$

Proof of Proposition 3. The proof is almost identical to that of Proposition 2. The difference is that the appropriate version of Lemma 2 is now much simpler to obtain. Assume $\Delta(\lambda) > 0$. Since $T(s, (\tilde{\theta}, \tilde{1})) \neq s$ for some s , then we immediately obtain that there are two states s, s' such that $\Pr(H_{s,s'}) = O(\beta)$ and $|\pi(H_{s'}^0) - \pi(H_s^0)| \geq \Delta(\lambda)$, which further implies, using the same argument as in Lemma 2 that:

$$|\pi(H_{s,s'}) - \pi(H_s^0)| \geq c\Delta(\lambda)$$

for some constant c .

7 Bibliography

Baumol, William J. and Richard E. Quandt (1964), "Rules of Thumb and Optimally Imperfect Decisions," *American Economic Review* 54(2), pp 23-46.

Benabou, R. and J. Tirole (2002), "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics*, 117(3), 871-915.

Benabou, R. and J. Tirole (2004), "Willpower and Personal Rules," *Journal of Political Economy*, 112 (4), 848-886.

Charness, Gary and Dan Levin (2005), "When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect," *American Economic Review*, 96, 1300-1309.

- Compte, O., and Andrew Postlewaite (2004), "Confidence-Enhanced Performance," *American Economic Review*, 94, 1536-1557.
- Compte, O. and A. Postlewaite (2009a), "Effecting Cooperation," mimeo, University of Pennsylvania.
- Compte, O. and A. Postlewaite (2009b), "Repeated Relationships with Limits on Information Processing", mimeo, University of Pennsylvania.
- Compte, O. and A. Postlewaite (2010a), "Plausible Cooperation," mimeo, University of Pennsylvania.
- Compte, O. and A. Postlewaite (2010b), "Mental Processes and Decision Making," mimeo, University of Pennsylvania.
- Cover, T. and M. Hellman (1970), "Learning with Finite Memory," *Annals of Mathematical Statistics* 41, 765-782.
- Dow, J. (1991), "Search Decisions with Limited Memory," *Review of Economic Studies* 58, 1-14.
- Festinger, L. (1957), *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Gilovich, T. (1991), *How We Know What Isn't So*. New York, NY: The Free Press.
- Griffin, D. and A. Tversky (1992), "The Weighing of Evidence and the Determinants of Confidence," *Cognitive Psychology* 24, 411-435.
- Hellman, M. E. and T. M. Cover (1973), "A Review of Recent Results on Learning with Finite Memory," *Problems of Control and Information Theory*, 221-227.
- Hvide, H. (2002), "Pragmatic beliefs and overconfidence," *Journal of Economic Behavior and Organization*, 2002, 48, 15-28.
- Jenkins, H.M. and W.C. Ward (1965), "Judgement of contingency between responses and outcomes," *Psychological Monographs*
- Josephs, R. A., Larrick, R. P., Steele, C. M., and Nisbett, R. E. (1992), "Protecting the Self from the Negative Consequences of Risky Decisions," *Journal of Personality and Social Psychology*, 62, 26-37.
- Kahneman, D., P. Slovic, and A. Tversky (1982), *Judgement under Uncertainty: Heuristics and Biases*, New York, NY: Cambridge University Press.
- Lamb, Emmet J. and Sue Leurgans (1979), "Does Adoption Affect Subsequent Fertility?", *American Journal of Obstetrics and Gynecology* 134, 138-144.
- Lichtenstein, S. and Fischhoff, B. (1977), "Do those who know more also know more about how much they know? The calibration of probability judgements," *Organizational Behavior and Human Performance*, 20, 159-183.
- Lipman, Barton (1995), "Information Processing and Bounded Rationality: A Survey," *The Canadian Journal of Economics* 28(1), pp. 42-67.
- Rosenthal, Robert W. (1993), "Rules of thumb in games," *Journal of Economic Behavior & Organization*, Elsevier, vol. 22(1), pages 1-13, September.
- Sedikides, C., Green, J. D., and Pinter, B. T. (2004), "Self-protective memory," In: D. Beike, J. Lampinen, and D. Behrend, eds., *The self and memory*. Philadelphia: Psychology Press.
- Ward, W. C. and H.M. Jenkins (1965), "The display of information and the judgment of contingency," *Canadian Journal of Psychology*

Wilson, A. (2004), "Bounded Memory and Biases in Information Processing," mimeo, University of Chicago.