Mental Processes and Decision Making*† PRELIMINARY AND INCOMPLETE

Olivier Compte
PSE, Paris

Andrew Postlewaite
University of Pennsylvania

March 2009; this version May 2012

*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 Yilmaz Kocer, 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.

†This paper is a revision of an earlier paper "Mental Processes and Decision Making".

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

The past half century has seen the integration of uncertainty into a wide variety of economic models which has led to increasingly sophisticated analysis of the behavior of agents facing uncertainty. The incorporation of uncertainty into an agent's decision making typically begins with the assumption that the agent has a probability distribution over the possible outcomes that would result from decisions she might take. It is standard to take the agent's beliefs as exogenously given. Savage (1954) is often given as justification for such modeling: when an agent's preferences over the random outcomes stemming from decisions the agent might take satisfy a set of seemingly plausible assumptions, the agent's choices will be as though the agent maximized expected utility. The agent's behavior will be exactly as if the agent had a utility function over outcomes and a probability distribution over

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 full[2], 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 that 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 economic models have at their core a decision-making module, and those models must somehow take account of agents' 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

[1] See Gilovich (1991), page 2.

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

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

[4] There 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).)

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 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

[5] Early 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.

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, 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 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:

$g(a,\theta)$	1	2
1	1	0
2	0	1

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).

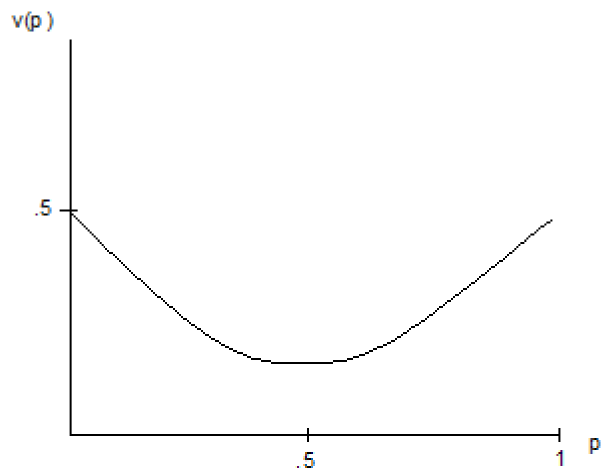


Figure 1: $v(\pi)$ when costs are distributed uniformly

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

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

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

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

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 $t = 0, 1, ...$, 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 $1 - \lambda$:

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

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.

Perceptions. We assume that the agents correctly interpret the signals they see. That is, when they see x, their perception is that x supports theory $\widetilde{\theta}(x)$ and that the strength of the evidence is $\widetilde{l}(x)$, and we assume that

$$\widetilde{\theta}(x) = \theta(x) \text{ and } \widetilde{l}(x) = l(x).$$

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 $\widetilde{l}$ associated with x. We denote by T the transition function:

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

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 $(\widetilde{\theta}, \widetilde{l})$. The event $A_0^+ = \{\widetilde{\theta} = 1\}$ corresponds to evidence in favor of state $\theta = 1$, while $A_0^- = \{\widetilde{\theta} \neq 1\}$ corresponds to evidence against $\theta = 1$. Transitions are as follows:

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

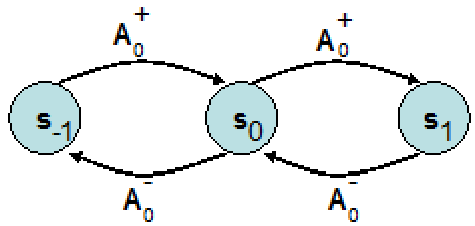


Figure 2: Transition function

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

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. For each given state θ**,** 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, $\phi_{\theta}(.) \in \Delta(S)$. This distribution along with the probability distribution over the true state θ determines a joint distribution over (s, θ), denoted $\phi(.,.) \in \Delta(S \times \Theta)$, as well as a marginal distribution over states $\phi(.) \in \Delta(S)$.

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

[7] Papers 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).

[8] It 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.

that is, that the agent can maximize:

$$\sum_{s,\theta} \phi(s,\theta) E_c u(a(s,c),\theta,c).$$

Call $\pi(s) = \Pr\{\theta = 1 \mid s\}$ the Bayesian updated belief. For convenience, we will sometimes use the notation $\pi_\theta(s) = \Pr\{\theta \mid s\}$. We have $\pi_\theta(s) = \phi(s,\theta)/\phi(s)$, and the expected utility above can be rewritten as:

$$\sum_s \phi(s) \sum_\theta \pi_\theta(s) E_c u(a(s,c),\theta,c).$$

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

$$v(S,T) = \sum_s \phi(s) v(\pi(s)).$$

Note that while one can compute, as a modeller, the distributions ϕ_θ and posterior beliefs π, we do not assume that the agent knows them. Rather, our assumption is that the agent can identify optimal behavior and thus his behavior coincides with that of an agent who would compute posteriors and behave optimally based on these posteriors.

We illustrate our approach next with specific examples.

Computations. We consider the mental process of example 1, and we illustrate how one computes the distribution over states prior to decision making. Define

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

Thus p_1 for example corresponds to the probability that the agent correctly perceives the true state as being $\theta = 1$ when the actual state is $\theta = 1$. We represent a distribution ϕ over states as a column vector:

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

and we let ϕ^0 denote the initial distribution over states of mind (i.e., that distribution 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 of example 1.

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,

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

These expressions can then be used to derive $\phi(s,\theta)$, $\phi(s)$ and $\pi_\theta(s)$.[9]

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 distributions over states ϕ_θ and further, expected welfare $v(S,T)$.

A fully symmetric case.

We illustrate our ideas under a symmetry assumption. We assume that $\pi_0 = 1/2$ and consider a symmetric signal structure:

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

Figure 3 below shows an example of such density functions for each of the two states $\theta = 1, 2$, assuming that $f(x \mid \theta = 1) = 2x$.[10]

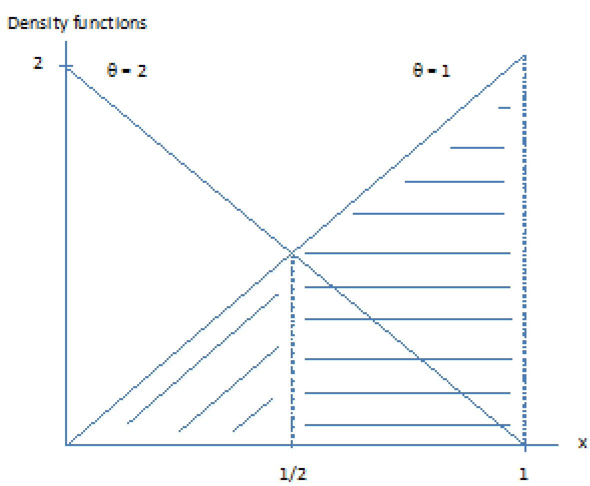


Figure 3: State contingent density functions

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

[9] $\phi(s,1) = \pi^0\phi_1(s)$, $\phi(s,2) = (1-\pi^0)\phi_1(s)$, and $\phi(s) = \sum_\theta \phi(s,\theta)$

[10] Given our symmetry assumption, this implies $f(x \mid \theta = 2) = 2(1-x)$. 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.

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 a "correct" signal is thus

$$p \equiv \int_{1/2}^{1} f(x \mid 1)dx = 3/4,$$

while the probability of a "misleading" or "wrong" signal is $1/4$. In other words, 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.

More generally, under assumption 1, one can define

$$p \equiv \Pr\{\widetilde{\theta} = 1 \mid \theta = 1\} = \Pr\{\widetilde{\theta} = 2 \mid \theta = 2\}$$

as the probability of a correct signal. The probability of a wrong signal is $1-p$, and the parameter p thus fully determines the transition probabilities across states, hence contingent on the true state θ the probability distributions ϕ_θ over the states (s_{-1}, s_0, s_1) are fully determined by p and λ. We have:

> **Proposition 1**: For $\theta = 1$, the probability distribution ϕ_θ over the mental states (s_{-1}, s_0, s_1) is: $\phi_\theta = ((1-p)\rho(1-\lambda p), \lambda(2-\lambda)\rho p(1-p), p\rho(1-\lambda(1-p)))$ where $\rho = \frac{1}{1-\lambda^2 p(1-p)}$

ϕ_2 is obtained by symmetry. In the case of Figure 3 for example, $p = 3/4$. So if the expected number of signals is very large (that is, if λ is close to 1), the probability distribution over his mental states (s_{-1}, s_0, s_1) is close to $(1/13,\ 3/13,\ 9/13)$ if the true state is $\theta = 1$, and close to $(9/13,\ 3/13,\ 1/13)$ if the true state is $\theta = 2$.

These distributions illustrate how the true state θ affects the mental state s in which the agent is likely to be in upon taking a decision. Because signals are not perfect, the agent's mental state is not a perfect indication to the decision maker of the true state, but it is an *unbiased* one in the following minimal sense:

$$\begin{aligned} \phi_\theta(s_1) &> \phi_\theta(s_{-1}) \text{ when } \theta = 1 \\ \phi_\theta(s_1) &< \phi_\theta(s_{-1}) \text{ when } \theta = 2 \end{aligned}$$

Finally, proposition 1 permits to derive the expected distribution ϕ over states,[11] and thus the posterior beliefs at each mental state. It also 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:

$$2\phi(s_1)v(\pi(s_1)) + (1 - 2\phi(s_1))v(\frac{1}{2}),$$

[11] $\phi = \frac{1}{2}(\phi_1 + \phi_2)$

where $\pi(s_1)$ and $\phi(s_1)$ are derived from Proposition 1.[12]

As one expects, welfare increases with the precision of the signal (p): as p or λ get closer to 1, bayesian 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.

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:

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

In other words, the event $A = \{\widetilde{l} < 1 + \beta\}$ 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, the agent's welfare contingent on being in state s_1 or s_{-1} will be higher (posterior 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 $f(x \mid \theta = 1) = 2x$. 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.

[12] We have $\pi(s_1) = p\frac{1-\lambda(1-p)}{1-2\lambda p(1-p)}$ and $\phi(s_1) = \frac{1-2\lambda(1-p)p}{2-2\lambda^2(1-p)p}$. Note that $\pi(s_1) > p$ for all values of λ. Intuitively, being in 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 of 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).

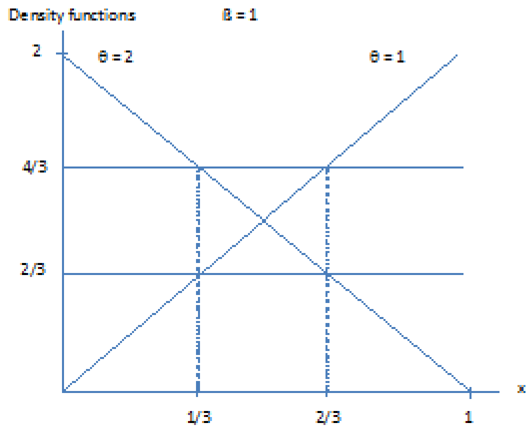


Figure 4

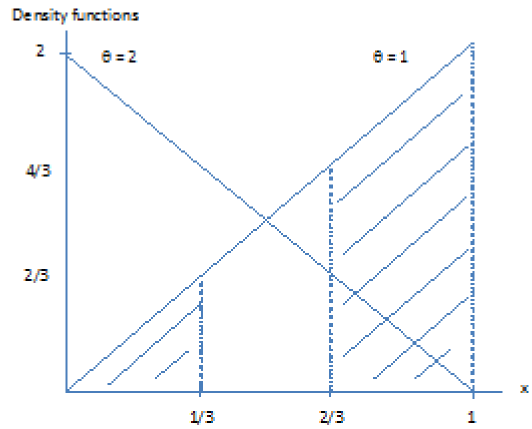


Figure 5

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. In addition, posterior beliefs are more accurate: $\pi(s_1) = 25/26$ under T^0, and $\pi(s_1) = 9/10$ under T^0.

Larger β would lead to even more accurate posteriors, with $\pi(s_1)$ converging to 1 as β goes to infinity. But this increase in accuracy comes at a cost. The probabilities $\phi(s)$ are those associated with the ergodic distribution, but the decision maker will get a finite (random) number signals, and the expected number of signals 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 leads 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

$y = \Pr\{\widetilde{\theta} = 1, \widetilde{l} < 1 + \beta \mid \theta = 1\}$ and $z = \Pr\{\widetilde{\theta} = 2, \widetilde{l} < 1 + \beta \mid \theta = 1\}$.[13] 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 the distributions ϕ_θ, the posterior belief $\pi(s_1)$ and the probability $\phi(s_1)$, hence the welfare associated with T^β. Increasing β typically raises $\pi(s_1)$ (which is good for welfare), but, for large values of β, it also 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.

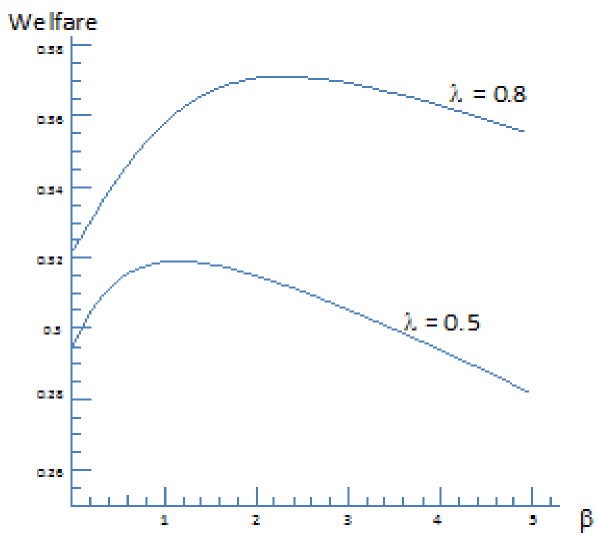


Figure 6: Welfare as a function of β

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

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

0 would not be optimal either. The graph illustrates the basic trade-off: large $\beta' s$ run the risk of never leaving the initial state while small $\beta' s$ 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 $\widetilde{\theta} \in \{1, 2\}$ a transition to state $T^0(s, \widetilde{\theta})$. 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 $\{\widetilde{l} > 1 + \beta\}$, and that does not generate a change in the agent's mental state when $\{\widetilde{l} < 1 + \beta\}$.

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.[14] 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.[15]

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

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

[14] $\underline{W} = v(\pi_0)$.

[15] In 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.

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.

3.3 Generalization

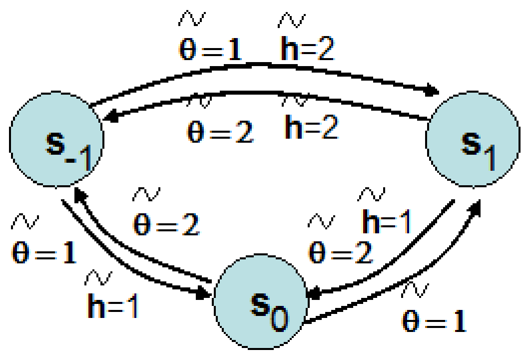
A simple mental process does not distinguish mild and strong evidence. As we shall see, a more sophisticated mental process that distinguishes between mild and strong evidence can sometimes improve welfare. We next extend the class of mental systems we consider to include a class of sophisticated mental processes and show that our insight that ignoring weak evidence improves welfare holds in this larger class.

Let us emphasize that our objective is not to characterize optimal mental processes. A mental process that is optimal for a particular problem would need to be tailored to the particular distribution f, the parameter λ for that problem as well as the particular utilities attached to each decision. Rather, our view is that a mental process will be applied to many problems, and that in general, identifying how a mental process should be amended to improve welfare, or identifying the class of problems for which one mental process would outperform another are difficult tasks. Our result illustrates that there is, however, one direction of change, ignoring weak evidence, that unambigously improves welfare. **O: I'm happy to leave the previous sentence as it is, but we should understand that it may not be precisely correct. The set of problems that a mental process is meant to apply to may be such that for every $\beta>0$, there is SOME problem in the class for which the amended process that ignores weak information performs worse than the unamended process that does not ignore weak evidence. Furthermore, the probabilty distribution over problems might be such that expected welfare is maximized at $\beta = 0$.**

A class of sophisticated mental system.

We first define a class of *level-k transitions.* Consider an increasing sequence of $(\beta_1, ..., \beta_{k-1})$, and set $\beta_0 = 0, \beta_k = +\infty$. Any perception $\widetilde{l} \in (1+\beta_{\widetilde{h}-1}, 1+\beta_{\widetilde{h}})$ with $\widetilde{h} \in \{1, ..., k\}$ is labelled as a perception of strength $\widetilde{h}$. 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 $(\widetilde{\theta}, \widetilde{h})$.

The figure below provides an example of a level 2 mental system, where strong signals ($\widetilde{h} = 2$) never induce a transition to state s_0.



Given any level k mental system for which $T(s, (\widetilde{\theta}, \widetilde{1})) \neq s$ for some s, one may consider a modified mental system T^{β} that would coincide with T when $\widetilde{l} > 1 + \beta$, but that would ignore signals for which $\widetilde{l} < 1 + \beta$.

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 (S, T) such that $T(s, (\widetilde{\theta}, \widetilde{1})) \neq s$ 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 $\underline{q}(\lambda) = \min_{s \in S} \phi(s)$ denote minimum weight on any given state when (S, T) is used.
>
> Proof: See Appendix.

A welfare comparison.

To conclude this section, we illustrate how a sophisticated mental process may lead to higher welfare, and why it need not.

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

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

Finally, we let ν denote the probability of a strong signal.[17] The three numbers $(\nu, \bar{l}, \underline{l})$ fully characterize the distribution over signals.[18]

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.

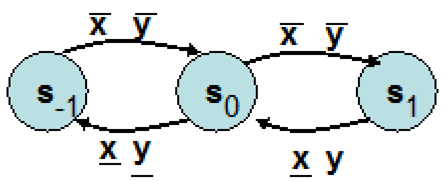


Figure 7: Simple mental process

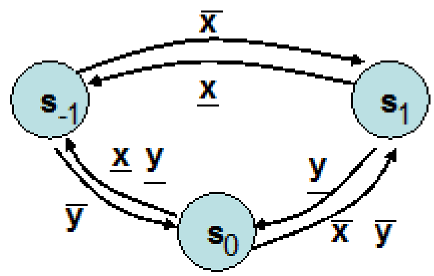


Figure 8: Sophisticated mental process

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 **of** $\bar{l}$ 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.

[17] This probability is assumed to be independant of the state θ.

[18] For example $\Pr\{\bar{x} \mid 1\} = \frac{\bar{l}\nu}{1+\bar{l}}$, and $\Pr\{\bar{y} \mid 1\} = \frac{\underline{l}(1-\nu)}{1+\underline{l}}$.

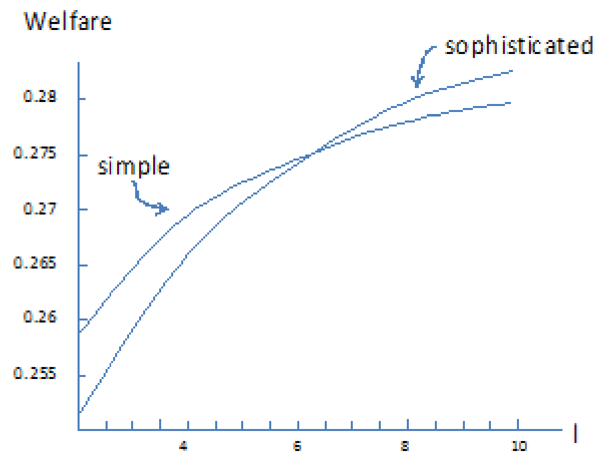


Figure 9: Simple vs. sophisticated transition

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.

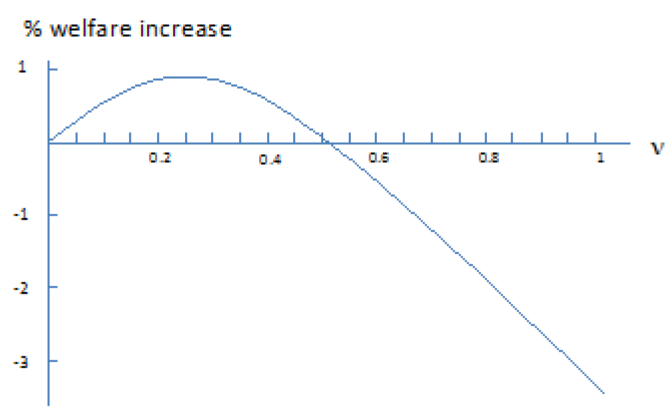


Figure 10: Relative gain with sophisticated transition

These graphs show that there may be an advantage to more sophisticated transitions, though not always.

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.[19]

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 $\theta = -1$ 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.

[19] Note that we do not have in mind that the decision maker would have biased posterior beliefs. We have asssumed throughout the paper that the decision maker is able to maximize 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.

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 $f(. \mid \theta_{-1})$ being the uniform distribution on $[0, 1]$.

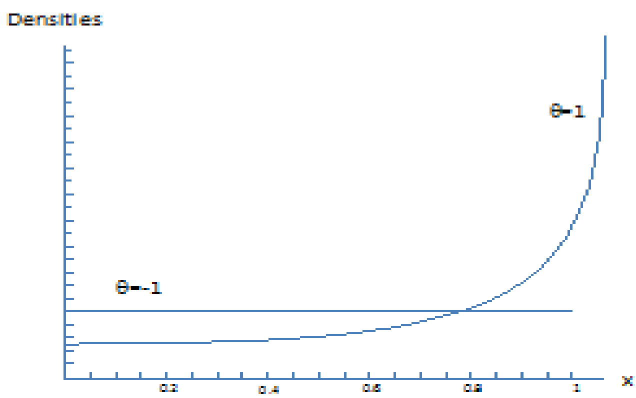


Figure 11: An asymmetric example

Each signal x supports theory $\widetilde{\theta} \in \{\theta_1, \theta_{-1}\}$ depending on whether $f(x \mid \theta_1)$ or $f(x \mid \theta_{-1})$ 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.