Uncommon Priors Require Origin Disputes

Robin Hanson
Department of Economics
George Mason University*

July 2006, First Version June 2001

Abstract

In standard belief models, priors are always common knowledge. This prevents such models from representing agents' probabilistic beliefs about the origins of their priors. By embedding standard models in a larger standard model, however, *pre-priors* can describe such beliefs. When an agent's prior and pre-prior are mutually consistent, he must believe that his prior would only have been different in situations where relevant event chances were different, but that variations in other agents' priors are otherwise completely unrelated to which events are how likely. Due to this, Bayesians who agree enough about the origins of their priors must have the same priors.

Introduction

The most standard way to model a set of agents with beliefs that change over time is to describe a set of possible states, a prior probability distribution for each agent, and an information partition for each agent at each time. In some such standard models, all agents have the same prior, while in other models priors differ. Some argue that common priors are implied by common knowledge of rationality, since rational agents with the same information should have the same beliefs (Aumann, 1998). Others do not find this argument compelling (Morris, 1995; Gul, 1998).

Since it is logically possible either that agent priors are common, or that they are uncommon, one might imagine that an agent could be uncertain about the priors of other agents. Within a standard model, however, no agent can be unsure of any agent's prior; every prior is always common knowledge. This is problematic, however, as it seems to preclude agents from using probabilities to reason about the origins of their priors.

For example, if there were such a thing as a gene for optimism versus pessimism, you might believe that you had an equal chance of inheriting your mother's optimism gene or your

* *rhanson@gmu.edu* http://hanson.gmu.edu 703-993-2326 FAX: 703-993-2323 MS 1D3, Carow Hall, Fairfax VA 22030

father's pessimism gene. You might further believe that your sister had the same chances as you, but via an independent draw, and following Mendel's rules of inheritance. You might even believe that humankind would have evolved to be more pessimistic, had they evolved in harsher environments. Beliefs of this sort seem central to scientific discussions about the origin of human beliefs, such as occur in evolutionary psychology.

Beliefs about the origins of priors also seem relevant to the rationality of priors. For example, if you learned that your strong conviction that fleas sing was the result of an experiment, which physically adjusted people's brains to give them odd beliefs, you might well think it irrational to retain that belief (Talbott, 1990). Similarly it might be irrational to be more optimistic than your sister simply because of a random genetic lottery.

This paper presents a theoretical framework in which agents can hold probabilistic beliefs about the origins of their priors, and uses this framework to consider how such beliefs might constrain the rationality of priors. The basic approach is to embed a set of standard models within a larger encompassing standard model. Each embedded model differs only in which agents have which priors, while the larger encompassing model includes beliefs about which possible prior combinations might be realized.

Just as beliefs in a standard model depends on ordinary priors, beliefs in the larger model depend on *pre-priors*. We do not require that these pre-priors be common; pre-priors can vary. But to keep priors and pre-priors as consistent as possible with each other, we impose a *pre-rationality* condition. This condition in essence requires that each agent's ordinary prior be obtained by updating his pre-prior on the fact that nature assigned the agents certain particular priors.

This pre-rationality condition has strong implications regarding the rationality of uncommon priors. Consider, for example, two astronomers who disagree about whether the universe is open (and infinite) or closed (and finite). Assume that they are both aware of the same relevant cosmological data, and that they try to be Bayesians, and therefore want to attribute their difference of opinion to differing priors about the size of the universe.

This paper shows that neither astronomer can believe that, regardless of the size of the universe, nature was equally likely to have switched their priors. Each astronomer must instead believe that his prior would only have favored a smaller universe in situations where a smaller universe was actually more likely. Furthermore, he must believe that the other astronomer's prior would not track the actual size of the universe in this way; other priors can only track universe size indirectly, by tracking his prior. Thus each person must believe that prior origination processes make his prior more correlated with reality than others' priors.

As a result, these astronomers cannot believe that their differing priors arose due to the expression of differing genes inherited from their parents in the usual way. After all, the usual rules of genetic inheritance treat the two astronomers symmetrically, and do not produce individual genetic variations that are correlated with the size of the universe.

This paper thereby shows that agents who agree enough about the origins of their priors must have the same prior. This is a new argument for common priors, and one that depends only on consistency relations between the beliefs of a single agent.

Analysis

A standard[1] Bayesian belief model $M = (\Omega, p, \Pi)$ regarding N agents consists of a finite set Ω of states, an assignment $p = (p_1, p_2, \ldots, p_N)$ of a prior probability distribution p_i over Ω for each agent i, and an information history $\Pi = (\Pi^t)_{t \in T}$ over times T, where $\Pi^t = (\Pi^t_1, \Pi^t_2, \ldots, \Pi^t_N)$ and each Π^t_i is a partition[2] of Ω. At state $\omega \in \Omega$ and time $t \in T$, agent i assigns to event E the subjective probability $p_{it\omega}(E) = p_i(E|\Pi^t_i(\omega))$, where $p_i(A|B) \equiv p_i(A \cap B)/p_i(B)$ and $p_i(E) \equiv \sum_{\omega \in E} p_i(\omega)$.

In a standard model M, every prior p_i (a map between states and probabilities) is always common knowledge. To allow uncertainty about the priors of a standard model, let us introduce an extended model $\tilde{M} = (\Omega, \mathcal{P}, \Pi, q, \Gamma)$, where $\mathcal{P}$ is a set of possible priors p_i over Ω, and where q and Γ are assignments of priors and information partitions over an extended state space $\tilde{\Omega} = \Omega \times \mathcal{P}^N$.

A *pre-state* $\tilde{\omega} = (\omega, p) \in \tilde{\Omega}$ combines an ordinary state $\omega \in \Omega$ and a prior assignment $p = (p_1, p_2, \ldots, p_N) \in \mathcal{P}^N$. For each ordinary event $E \subset \Omega$ let us define a corresponding event $\tilde{E} \equiv \{(\omega, p) : \omega \in E\} \subset \tilde{\Omega}$ in the extended space, and for each ordinary information set $\Pi^t_i(\omega)$, let us define a corresponding information set

$$\tilde{\Pi}^t_i((\omega, p)) \equiv \{(\omega', p') : p = p', \omega' \in \Pi^t_i(\omega)\}$$

in the extended space. Let us also project an ordinary prior p_i into a prior $\tilde{p}_i$ over the extended space by defining

$$\tilde{p}_i(\tilde{E}|p) \equiv p_i(E), \tag{1}$$

where prior assignment p treated as an event denotes $p = \{(\omega, p') : p' = p\}$. Agent prior p_i can also be treated as the event $p_i = \{(\omega, (p'_1, p'_2, \ldots, p'_i, \ldots, p'_N)) : p'_i = p_i\}$, where $\bigcap_i p_i = p$.

These definitions let us see the extended space $\tilde{\Omega}$ as being partitioned into many ordinary models $M(p) = (p, \tilde{p}, \tilde{\Pi})$, which are identical to each other and to our original model M, except for having differing prior assignments p. Models $M(p)$ describe the belief history of N ordinary agents i at times $t \in T$. Within each model $M(p)$, the beliefs of every ordinary agent i at time t are described in terms of this extended space $\tilde{\Omega}$ using the extended prior $\tilde{p}_i$ and the extended information $\tilde{\Pi}^t_i$.

Even in this extended space, the priors p are always common knowledge for agents i. This fact prevents any agent i from using beliefs derived from his prior $\tilde{p}_i$ and information $\tilde{\Pi}^t_i$ to express beliefs about the origins of his and others' priors. To allow the expression of such beliefs about origins, let us introduce the concept of a *pre-agent* i.

We introduce pre-agent i as a way to represent certain counter-factual beliefs of agent i about the origins of priors. These beliefs are counter-factual in the sense that they are counter to the fact that agent i must always know everyone's priors. While we want the beliefs of agent i and pre-agent i to agree as much as possible, some of pre-agent i's beliefs will describe what he considers to be reasonable beliefs about the processes that produced the prior of agent i and the other agents. Such beliefs about prior origins might be based on beliefs about genetic processes, cultural processes, or any other processes that the agent believed to be relevant in producing the particular priors that the agents were given.

Formally, each pre-agent i has a pre-prior q_i and pre-information Γ_i^t over the extended space $\tilde{\Omega}$ for times $t \in S$. The key to allowing uncertainty about agent priors is to introduce new earlier times ($s \in S$ such that $s < t$ for all $t \in T$) when agent i has no beliefs. At such times s, pre-agent i can be uncertain about many aspects of the process that produced the various agents' priors, including which exact priors p resulted. For this to be possible, the pre-information Γ_i^s of pre-agent i should not make the event p common knowledge among pre-agents at early times s.

Since pre-agent i represents the counterfactual beliefs of agent i, the beliefs of agent i and pre-agent i should agree as much as possible. Therefore at the times when both agents and pre-agents have beliefs, agent and pre-agent priors and information partitions should agree. That is, for all i and all $t \in T \cap S$, we should have $\Gamma_i^t = \tilde{\Pi}_i^t$.

Furthermore, priors p_i and pre-priors q_i should satisfy a *pre-rationality* condition,

$$q_i(\tilde{E}|p) = \tilde{p}_i(\tilde{E}|p). \tag{2}$$

This condition says that agent priors are as if agent i had acquired his prior p_i by conditioning his pre-prior q_i on the fact that nature assigned him and others certain priors. Note that we do *not* assume pre-priors are common. Our assumptions do not directly constrain the relations between agent beliefs, or the relations between pre-agent beliefs, but only the relations between pre-agent and agent beliefs.

Equations 1 and 2 combine to give

$$q_i(\tilde{E}|p_1, p_2, p_3, \ldots, p_i, \ldots, p_N) = p_i(E). \tag{3}$$

The following results all follow trivially from equation 3.

Theorem 1 *In pre-prior q_i, given prior p_i any event $\tilde{E}$ is independent*[3] *of other priors $p_{j\neq i}$.*

That is, once pre-agent i knows prior p_i, no other prior p_j can be informative about any event E, nor can any event E be informative about any other prior p_j.

Theorem 2 *In pre-prior q_i, any $\tilde{E}$ is directly dependent on prior p_i via $q_i(\tilde{E}|p_i) = p_i(E)$.*

To pre-agent i, the prior p_i and any event E are always informative about each other.

Corollary 1 *If $q_i(\tilde{E}|\, p_i = P) = q_i(\tilde{E}|\, p_i = P')$, then $P(E) = P'(E)$.*

My prior would only change when the events it forecasts become more or less likely.

Corollary 2 *If $q_i(\tilde{E}|\, p_i = P,\, p_j = P') = q_i(\tilde{E}|\, p_i = P',\, p_j = P)$, then $P(E) = P'(E)$.*

If event E were just as likely in situations where my prior had been exchanged with someone else's prior, those priors must be the same regarding event E. And via Bayes' rule, the same holds if exchanged priors are just as likely given E, and just as likely given not E.

One concern about the above analysis is that our beliefs about the processes that produce priors are based on information we have received during our lives as agents. So this information would not have been available to our pre-agents, who we imagine held beliefs before we ever had any beliefs.

Fortunately, we can generalize equation 3 to include any background information B,

$$q_i(\tilde{E}|p_1, p_2, p_3, \ldots, p_i, \ldots, p_N, \tilde{B}) = p_i(E|B).$$

Thus theorems 1 and 2 and corollaries 1 and 2 all easily generalize to condition on B. For example, theorem 2 becomes $q_i(\tilde{E}|p_i\tilde{B}) = p_i(E|B)$. So if we set B to represent the background information on which we base our beliefs about the basic processes that produce priors, all of the above results will apply conditional on that background information.

Discussion

These constraints on beliefs about the origins of priors are strong and highly asymmetric. Each agent must believe that his prior would "track truth" in the sense that his prior would only assign a higher probability to an event in situations where that event actually *was* more likely. Furthermore, he must believe that other agent's priors would only track truth to the extent that their priors covaried with his prior; he believes any additional variation in the priors of others must be completely unrelated to any other events of interest.

In contrast, standard scientific beliefs about the origins of individual human variations do not offer much support for the belief that some people's initial beliefs tendencies track truth much better than other people's tendencies.

For example, some have argued that many general attitudes, such as general optimism or pessimism, are influenced by our genes (Olson, Vernon, Harris, & Jang, 2001). (Others disagree with this claim.) Mendel's rules of genetic inheritance, however, are symmetric and random between siblings. If optimism were coded in genes, you would not acquire an optimism gene in situations where optimism was more appropriate, nor would your sister's attitude gene track truth any worse than your attitude gene does.

Thus it seems to be a violation of pre-rationality to, conditional on accepting Mendel's rules, allow one's prior to depend on individual variations in genetically-encoded attitudes. Having your prior depend on species-average genetic attitudes may not violate pre-rationality, but this would not justify differing priors within a species.

Similar problems apply to cultural processes that produce priors. If priors are transmitted culturally via children copying visible adults, standard theories about individual variations in such culturally-transmitted belief tendencies offer little support for the idea that some children are better able to select the most truth-tracking cultural elements from among the available cultural transmissions. Perhaps some children could better extract relevant information from the available cultural transmissions, but any differences in information must represented in differing information partitions, and not in differing priors.

Without some basis for believing that the process that produced your prior was substantially better at tracking truth than the process that produced other peoples' priors, you appear to have no basis for believing that beliefs based on your prior, are more accurate than beliefs based on other peoples' priors.

Conclusion

In standard models priors are common knowledge, which makes it hard to express probabilistic beliefs that agents might have about the origins of their priors. This paper has introduced a formal framework to allow standard agents to reason probabilistically about the origins of their priors. Standard models corresponding to different priors are embedded in a larger space where priors can vary, and pre-agents are imagined who can be uncertain about which priors ordinary agents will get. Pre-priors, the priors of these pre-agents, then describe counterfactual beliefs an agent has regarding the origins of ordinary priors.

Using this concept of a pre-prior, this note has presented a new argument for the rationality of common priors. The argument is that a simple plausible constraint relating rational priors and pre-priors seems to conflict with our usual scientific stories about the origins of individual variations in early belief tendencies. While people often disagree about the origins of humankind, they also seem to accept the idea that nature treated individuals symmetrically ex ante, and that most other topics of disagreement are irrelevant to estimating those origin processes. It thus seems hard to attribute most human disagreements to these sort of rationally differing priors.

Acknowledgments

For their comments, I thank Nick Bostrom, Tyler Cowen, Ronald Heiner, Sahotra Sarkar, William Talbott, one very helpful anonymous referee, and ten other anonymous referees. I thank the Center for Study of Public Choice, the Mercatus Center, and the International Foundation for Research in Experimental Economics for financial support.

Notes

[1] Belief hierarchies are also a standard form. The mapping between these forms is unique given common priors.

[2] It is common to also require that Π_i^t weakly refine Π_i^s when $t > s$.

[3] In distribution P, event A is independent of B when $P(A|B) = P(A)$. In P, event A is independent of B conditional on C when $P(A|BC) = P(A|C)$. Independence is a symmetric relation.

References

Aumann, R. J. (1998). Common Priors: A Reply to Gul. *Econometrica*, *66*(4), 929–938.

Gul, F. (1998). A Comment on Aumann's Bayesian View. *Econometrica*, *66*(4), 923–927.

Morris, S. (1995). The Common Prior Assumption in Economic Theory. *Economics and Philosophy*, *11*, 227–253.

Olson, J. M., Vernon, P. A., Harris, J. A., & Jang, K. (2001). The Heritability of Attitudes: A Study of Twins. *Journal of Personality and Social Psychology*, *80*(6), 845–860.

Talbott, W. (1990). *The Reliability of the Cognitive Mechanism: A Mechanist Account of Empirical Justification*. Garland Publishing, New York.