

Edenic Orgulity

GORDON BELOT

I. In the Garden

1. Suppose that we are enjoying the Garden of Eden one day when The Gardener happens by and makes the following speech.

I thought you might enjoy playing some games—keep you out of trouble.

I have three coins. The first is fair. The second has a bias in favour of heads given by a rational number (strictly) between zero and one, and the the third has a bias given by an irrational number between zero and one. I flipped the fair coin to choose between the second and third coins and dubbed the chosen one The Coin. Then I repeatedly tossed The Coin to generate an infinite binary sequence (o's corresponding to heads, 1's to tails), which I call The Sequence.

I have in mind two games. Each has the following structure. I pose a question about The Sequence. You propose a *strategy* for answering this question on the basis of finite data sets: a function that takes as input binary strings (encoding outcomes of finite numbers of tosses of The Coin) and gives as output conjectures (possible answers to the question posed). Let \hat{r}_1 be the conjecture that your strategy yields when it sees the first bit of The Sequence, \hat{r}_2 be the conjecture that it yields when when show the first two bits of The Sequence, and so on. Your strategy

succeeds if the sequence of conjectures $\hat{r}_1, \hat{r}_2, \hat{r}_3, \dots$ converges to the truth about The Sequence. If your strategy succeeds then you get something nice—a piece of that fruit you like. If your strategy fails then something unpleasant happens to you (not something infinitely bad—just very, very bad).

The first game. The question is: What is the limiting relative frequency R of o's in The Sequence?

The second game. The question is: Is R a rational number or an irrational number?

Should we play the first game? Should we play the second game?

II. What Rationality Permits

2. I myself do not enjoy unpleasantness. And I have been around enough to know that, when motivated, The Gardener can make things very unpleasant indeed. And, frankly, fruit is not all *that* hard to come by in the Garden of Eden. Knowing that winning one of these games has only small positive utility whereas losing one has unknown but great disutility, I am going to play only if I have a strategy that has no chance of losing no matter what the bias of The Coin. I hope the reader will agree that this attitude is rationally permitted.

Now, by the Strong Law of Large Numbers, no matter what the actual bias R of The Coin, there is chance one that tossing it an infinite number of times will generate an infinite sequence in which the relative frequency of heads is R .¹ It follows that I have no chance of losing the first game if I use the straight rule to generate my conjectures, guessing that the bias is k/n when I am told that heads has come up k times in n tosses.

There are no similarly good strategies for the second game—

Version of June 2023.
Earlier versions of this were presented in the Orgulity Octagon in Ann Arbor, at Butterjoint in Pittsburgh, and at a Sgt. Pepperoni's in Irvine. For helpful discussion, thanks very much to Jeff Barrett, Jim Joyce, Calum McNamara, Chris Mierzewski, Laura Ruetsche, and Francesca Zaffora Blando. Thanks also to two helpful (if, ultimately, implacable) anonymous referees.

¹See, e.g., Billingsley (1995), Theorem 6.1. Here and throughout we assume that since The Sequence is generated by repeated tossing of The Coin, we can treat its bits as independent and identically distributed random variables.

unsurprisingly, since, intuitively, we are getting little or no information about whether or not R is rational from the finite data sets that we see.² Indeed, it can be shown that every strategy for playing the second game has one or the other of the following features.³ (i) There is a nonempty open subinterval J of the unit interval such that for each rational number $b \in J$, there is a non-zero chance that tossing a coin of bias b would generate a sequence that would lead the strategy to lose the second game. (ii) There is a set K of irrational numbers in the unit interval that is dense and uncountable (indeed co-meagre) and such that for any $b \in K$, there is zero chance that tossing a coin of bias b would generate a sequence that would lead the strategy to win the second game. Thus there is no strategy that gives me zero chance of losing the second game. So I am not going to play.

3. Maybe you are sharper than I am and are able to estimate the disutility of losing these games—and so are able to determine a probability p such that it is worth playing these games, so long as the chance of winning is at least p . But this on its own will not allow you to determine whether it is worth playing the second game because you cannot calculate the chance of winning without knowing the chance, for each subinterval of the unit interval, that bias of The Coin lies in that subinterval. But we do not have access to that information since we don't know anything about how The Gardener selected the pair of coins from which The Coin was chosen.

III. What Bayesianism Permits

4. According to factory-model Bayesianism, a *Bayesian agent* is someone who maximizes subjective expected utility in choosing actions, whose credal state at any time is representable by a (countably additive) probability measure on the relevant space of possibilities, who updates

²Cover (1973) is the classic treatment of this problem. For further discussion and references, see Kleijn (2023), §9.1.2.

³See the proof of Theorem 2 in Koplowitz, Steif, and Nerman (1995).

these credal states by conditionalization on available data, and who defers to known chance facts (in the sense that their credal states agree with any known facts about objective chances). We will also henceforth restrict attention to Bayesian agents for whom declining to play either of our games has less utility than playing and winning (and for whom losing has finite disutility).

After hearing The Gardener's speech, the credences of a Bayesian agent concerning the bias of The Coin will be encoded in a probability measure μ_0 on the open unit interval (the space of possible biases of The Coin). Because μ_0 defers to known chance facts, we can assume that if σ is a binary string containing k 0's and ℓ 1's, then conditional on the bias of The Coin being r , μ_0 assigns probability $r^k(1-r)^\ell$ to the proposition $\llbracket\sigma\rrbracket$ that The Sequence begins with σ . This allows us to calculate, for each binary string σ , the probability that μ_0 assigns to $\llbracket\sigma\rrbracket$.⁴ It follows that μ_0 induces a probability measure $\tilde{\mu}_0$ on Cantor space (the space of infinite binary sequences).⁵ Then a Bayesian agent whose data is given by a k -bit binary string σ will have a posterior probability distribution $\tilde{\mu}_k = \tilde{\mu}_0(\cdot | \llbracket\sigma\rrbracket)$ on Cantor space, which will determine a unique probability measure μ_k over the possible biases of The Coin.⁶

Fix a Bayesian agent with a prior μ_0 on the unit interval that induces prior $\tilde{\mu}_0$ on Cantor space. For either of our games, and for any strategy π for playing that game, we can consider the set W_π of binary sequences that lead to victory when that strategy is played. Then $\tilde{\mu}_0(W_\pi)$ is the probability that our agent assigns to winning the game if strategy π is played. Our agent will certainly agree to play a game if there is a strategy for which this probability is one.

⁴This is the integral over the unit interval of $f(r) := r^k(1-r)^\ell$ (with respect to the measure μ_0).

⁵The Carathéodory Extension Theorem implies that fixing the probability that μ_0 assigns to each $\llbracket\sigma\rrbracket$ determines a unique probability measure $\tilde{\mu}_0$ on Cantor space (defined on the sigma-algebra generated by the $\llbracket\sigma\rrbracket$)—see, e.g., Billingsley (1995), Theorem 11.3.

⁶This follows from de Finetti's Representation Theorem—see, e.g., Billingsley (1995), Theorem 35.10.

Consider the following strategy that our agent might use to play the first game: after the k th bit of The Sequence is revealed, calculate the mean of μ_k and use that as your conjecture as to the bias of The Coin. According to any prior μ_0 the set of binary sequences that result in victory for this strategy has probability one.⁷

Consider the following strategy that our agent might use to play The Gardener's second game: after the k th bit of The Sequence is revealed, calculate the probability that μ_k assigns to the set of rational numbers in the unit interval; if this number is greater than one-half, conjecture that the bias of The Coin is rational, otherwise conjecture that it is irrational. Any prior will assign measure one to the set of sequences for which this strategy leads to victory.⁸

So any Bayesian agent (with utilities obeying the minimal constraints imposed above) will agree to play both of the games offered by The Gardener.

5. There is a challenge here. It appears that rationality at least permits playing the first game while declining to play the second. But this pattern of behaviour will not be adopted by any agent satisfying the strictures of the factory-model Bayesian analysis of rationality. Should philosophers with Bayesian inclinations argue that, contrary to appearances, the factory-model Bayesian account is nonetheless correct? Or should they revise the factory-model analysis in order to allow the existence of rational agents who elect to play the first game but not the second?

6. This challenge is a relative of one issued in Belot (2013). In that paper it was observed that problems like guessing whether an infinite binary data stream gives the expansion of a rational number or an expansion of an irrational number have the following features: on the one hand, any method for approaching this problem that is open-minded in a

⁷That is: Bayes' estimates for this problem are consistent in the Bayesian sense—see, e.g., Billingsley (1995), 475.

⁸This is a consequence of Doob's Martingale Convergence Theorem—see, e.g., Schervish and Seidenfeld (1990), Theorem 2.

certain sense fails to converge to the truth for a large (indeed co-meagre) subset of possible data-streams; on the other hand, every agent whose credal state is represented by a probability measure and who updates by conditionalization is subjectively certain that their beliefs will converge to the truth.

It is possible to overcome this earlier challenge if one is willing to revise factory-model Bayesianism: by employing frameworks other than Cantor space to model unlimited binary data streams or by countenancing imprecise priors or merely finitely additive priors, one can show the existence of generalized Bayesian agents who are not certain of success in situations in which failure is typical for the relevant problem.⁹

The challenge developed in the present paper differs in a couple ways from the earlier version: open-mindedness plays no role in the new version; and whereas in the original version the problem at hand required agents to distinguish between a countable subset of Cantor space and its complement, in the new version the problem involves distinguishing between two uncountable subsets of Cantor space.¹⁰ It would be surprising (but interesting) if there were any difficulties in adapting the revisionist approaches noted above to the present version of the orgulity objection.

The real point of this paper is to address the silent majority of Bayesians who did not think of the original orgulity argument as putting pressure on the factory-model account. Questions about learning are

⁹For alternatives to Cantor space, see Huttegger (2015, 2022). On imprecise priors, see Weatherson (2015). On what can be accomplished using merely finitely additive priors, see Elga (2016), Cisewski *et al.* (2018), Pomatto and Sandroni (2018), Nielsen and Stewart (2019), and Gong *et al.* (2020). On finite additivity: for reasons discussed in Belot (2023), I think that it is crucial for Bayesians that at least some rationally permitted priors should be computable—and that (in rich contexts) there are interesting obstructions to developing an account of computable merely finitely additive probability measures.

¹⁰There is also a difference in rhetoric: the topological distinction between meagre and co-meagre subsets of Cantor space plays a peripheral role in the present discussion. On the epistemological (ir)relevance of this distinction, see Belot (2013), Huttegger (2015), Cisewski *et al.* (2018), Nielsen and Stewart (2019), and Zaffora Blando (2023).

sometimes shrugged off as being mere variants of the problem of induction. It is less easy to dismiss a puzzle about decision and action (even a fanciful one).

IV. Morals?

7. We have seen that each Bayesian agent is subjectively certain that the strategies considered in Section III above will lead to victory in both of The Gardener's games. But of course subjective certainty can come apart from objective chance: a Bayesian agent whose prior is concentrated on the first third of the unit interval and who faces a data stream generated by flipping a fair coin is never going to give accurate estimates of that coin's bias; a Bayesian agent who is certain that The Coin is fair if it has a rational bias is not going to perform well in the second game if the coin has some other rational bias.

8. The Bayesian strategy considered above for the first game has the feature that for any bias that The Coin might have, flipping a coin of that bias has zero chance of generating a sequence that would lead to that strategy losing the game if and only if the prior used assigns positive probability to each non-empty open subinterval of the real line.¹¹ On the other hand, there is *no* prior with the feature that playing the Bayesian strategy considered above for the second game is guaranteed to lead to victory in this sense. Indeed, since we are restricting attention to priors that defer to known chance facts, given how The Coin was chosen we can assume that μ_0 assigns equal probability to the set of rational biases and to the set of irrational biases. So it can be written in the form $\mu_0 = \frac{1}{2}\mu_R + \frac{1}{2}\mu_I$, where μ_R and μ_I are probability measures on the unit interval, with μ_R assigning probability one to the subset of rational numbers and μ_I assigning probability one to the subset of irrational numbers. If μ_R assigns positive probability to each rational number in the unit interval, then for each rational bias that The Coin might have, there is chance one that the strategy under discussion will lead to

victory if The Sequence is generated by flipping a coin of that bias.¹² It then follows from the result mentioned in Section II that for any such prior, there is a large set of irrational biases that The Coin might have such that there is no chance that an agent playing this strategy would win the second game.

9. Bayesian agents cannot see that the second game is harder than the first because in practical deliberation they do not distinguish between the aspects of their credences that correspond to known facts about objective chance and those that do not—and because in situations in which known facts about objective chance are too scant to determine a probability for each event of interest, in order to be rational by Bayesian lights agents must supplement known chance facts by further agent-relative probabilities. Even outside of the magical context of the Garden of Eden, where rewards and punishments can be determined by the behaviour of a completed ω -sequence, it is natural for anyone who accepts the existence of objective chance facts to be wary of purported accounts of rationality that require one to ignore in one's practical deliberations the distinction between these two sorts of probabilities.

References

- Belot, G. (2013) "Bayesian Orgulity." *Philosophy of Science* 80 483–503.
- Belot, G. (2023) "Unprincipled." Forthcoming in *The Review of Symbolic Logic*.
- Billingsley, P. (1995) *Probability and Measure* (third edition). Wiley & Sons.
- Cisewski, J., J. Kadane, M. Schervish, T. Seidenfeld, and R. Stern (2018) "Standards for Modest Bayesian Credences." *Philosophy of Science* 85 53–78.
- Cover, T. (1973) "On Determining the Irrationality of the Mean of a Random Variable." *The Annals of Statistics* 1 862–871.
- Elga, A. (2016) "Bayesian Humility." *Philosophy of Science* 83 305–323.

¹¹This follows from Theorem 1 of Freedman (1963).

¹²On this point see, e.g., Kleijn (2023), Example 9.5.5

- Freedman, D. (1963) "On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case." *Annals of Mathematical Statistics* 34 1386–1403.
- Gong, R., J. Kadane, M. Schervish, T. Seidenfeld, and R. Stern (2020) "Deceptive Credences." *Ergo* 7 28.
- Huttegger, S. (2015) "Bayesian Convergence to the Truth and the Metaphysics of Possible Worlds." *Philosophy of Science* 82 587–601.
- Huttegger, S. (2022) "Rethinking Convergence to the Truth." *Journal of Philosophy* CXIX 380–403.
- Kleijn, B. J. K. (2023) *The Frequentist Theory of Bayesian Statistics*. Forthcoming from Springer.
- Koplowitz, J., J. Steif, and O. Nerman (1995) "On Cover's Consistent Estimator." *Scandinavian Journal of Statistics* 22 395–397.
- Nielsen, M. and R. Stewart (2019) "Obligation, Permission, and Bayesian Orgulity." *Ergo* 6 03.
- Pomatto, L. and A. Sandroni (2018) "An Axiomatic Theory of Inductive Inference." *Philosophy of Science* 85 293–315.
- Schervish, M. and T. Seidenfeld (1990) "An Approach to Consensus and Certainty with Increasing Evidence." *Journal of Statistical Planning and Inference* 25 401–414.
- Weatherson, B. (2015) "For Bayesians, Rational Modesty Requires Imprecision." *Ergo* 2 20.
- Zaffora Blando, F. (2023) "Pride and Probability." Unpublished.