LESS IS MORE FOR BAYESIANS, TOO

GREGORY WHEELER

FRANKFURT SCHOOL OF FINANCE & MANAGEMENT

g.wheeler@fs.de

Lore has it that a fundamental principle of Bayesian rationality is for decision makers to never turn down the offer of free information. Cost-free information can only help you, never hurt you, and in the worst case will leave you at status quo ante. Purported exceptions to this principle are no exceptions at all, but instead involve a hidden cost to learning. Make those costs plain and the problem you face is one of balancing the quality of a choice against the costs to you of carrying it out, a trade-off that Bayesian methods are ideally suited to solve.

This piece of Bayesian lore, that rationality compels you to never turn down free information, is sometimes called *Good's Principle*, after I. J. Good's concise formalization of the reasoning behind it (Good 1967). But the argument goes back to the beginning of modern Bayesian probability theory, with remarks by (Ramsey 1931), an argument by (Savage 1972), the formalization of a key piece of it by (Raiffa and Schlaiffer 1961), followed thereafter by assertions in textbooks, starting with (Lindley 1965). Put a bit more carefully, Good's principle recommends to delay making a terminal decision between alternative courses of action if there is an opportunity to learn, at zero cost, the outcome of an experiment relevant to the decision (Pedersen and Wheeler 2015). This will put more carefully still in Section 2.

Objections to Good's principle have surfaced in the last half century, some of which are well known by now but others less so, forming part of a rich discussion of the value of information to rational decision making (Gigerenzer and Brighton 2009; Wakker 1988; Machina 1989; Seidenfeld 1994; Grünwald and Halpern 2004; Siniscalchi 2011; Hill 2013; Pedersen and Wheeler 2015). Since then a picture has emerged about the value of information that is more restricted and more nuanced than Good's principle suggests, suggesting a revision to Bayesian lore. For even in highly idealized settings, ignorance can be a virtue. Sometimes less is more for Bayesians, too.

1   ASYMMETRIC INFORMATION IN STRATEGIC GAMES

The first dent to this folk lore comes from the theory of games, where some strategic interactions can result in a player being better off having less information. George Akerlof's study of market failures created by asymmetric information is a classic example (Akerlof 1970). *Adverse selection* occurs when one side of a trade has less information than the other side and withdraws from trading from fear of being unfairly taken advantaged of by the more informed party. Akerlof offered the used car market as an example where adverse selection occurs, a particularly apt example in 1970. A used car salesmen will know which cars on the lot are bad and which are good, knowledge an ordinary consumer will not have. But the consumer will know that the dealer knows which car is of which quality type and recognize the upper hand the dealer has in any trade. Afraid of paying a good-car price and driving home in a bad one, the customer may choose to not buy any car at all. The reasoning for

this idealized single transaction generalizes, resulting in a market failure for used cars where nobody is willing to pay more than the going rate for a bad car.

Used car dealers have overcome their adverse selection problem by certifying the quality of used cars, and backing those claims with a warranty, thereby leveling the information playing field between dealers and costumers by letting costumers in on what dealers know about the quality of the cars they sell. (Making better cars has helped, too.) Yet, since the problem here is asymmetrical information, this isn't the only way to restore the market. Rather than making consumers as informed as dealers, another option is to make dealers as ignorant as consumers.

The following example, due to Martin Osborne, illustrates the ignorance option (Osborne 2003, 9.3). Imagine there are two states of the world, ω_1 and ω_2—which could be understood to correspond to the state in which a car is more likely to be good than bad and vice versa, for instance. Suppose there are two Bayesians, Player 1 and Player 2, but neither player knows which state of the world they are in. Both are ignorant, so they both assign a probability of one-half to ω_1 and one-half to ω_2. Figure 1 gives the payoff tables for Player 1 and Player 2, where the material difference to each player from their uncertainty over which state they are in, ω_1 or ω_2, is reflected in the last two columns of the respective payoff tables.

		Player 2		
		L	M	R
Player 1	T	$(1,2\varepsilon)$	$(1,0)$	$(1,3\varepsilon)$
	B	$(2,2)$	$(0,0)$	$(0,3)$

State ω_1

		Player 2		
		L	M	R
Player 1	T	$(1,2\varepsilon)$	$(1,3\varepsilon)$	$(1,0)$
	B	$(2,2)$	$(0,3)$	$(0,0)$

State ω_2

Figure 1: Payoffs to Players 1 and 2 in states ω_1 and ω_2 with $0 \le \varepsilon \le \frac{1}{2}$ (Osborne 2003).

Given this setup, with both players ignorant of the state, the strategy L is Player 2's unique best response to every strategy of Player 1, which yields Player 2 the expected payoff of $2 - 2(1-\varepsilon)p$, whereas M and R both yield $\frac{3}{2} - \frac{3}{2}(1-\varepsilon)p$ where p is the probability Player 1 assigns to T. Player 1's unique best respond to L is B. Therefore, (B,L) is the unique Nash equilibrium of the game, yielding each player a payoff of 2.

Now imagine that instead of both players being ignorant of which state they are in, exactly one of the players is informed of the state. Specifically, suppose Player 2 is informed of the state whereas Player 1 remains ignorant but nevertheless knows that Player 2 is informed. In this game (T, (R, M)) is the unique Nash equilibrium yielding to her at most a payoff of 1.5. Why? Choosing R is Player 2's best response in state ω_1 and her worst response in ω_2. Similarly, M is Player 2's best response in ω_2 and worst in ω_1. Player 1 knows this too, knows that Player 2 is informed of the state, thus knows that Player 2 will never choose L. With column L removed from consideration, Player 1's best response to (M,R) is T.

Despite her information advantage over Player 1, Player 2's payoff in this second game is 3ε in each state, which is at most 1.5. Thus, Player 2 is worse off learning the state than remaining ignorant. Given the choice between the original game, where both players are ignorant, and the second game in which Player 1 remains ignorant but Player 2 is informed, it is rational for Player 2 to choose to remain ignorant, even if the information about the state

2

is offered to her for free.

Akerlof's and Osborne's examples are part of a broader collection of counter-intuitive results that can arise when the rational choice of one player changes the probability assessments of another about which state will occur. In this case, the negative value of information stems from *act-state dependence* of Player 1's strategic response to Player 2's informed choice. Osborne seems to think that the prospect of information having negative value appears only in games, not in decision problems:

> A decision-maker in a single-person decision problem cannot be worse off if she has more information: if she wishes, she can ignore the information. In a game the same is not true: if a player has more information and the other players know that she has more information, then she may be worse off (Osborne 2003, p. 281).

This position that Osborne expresses, that Good's principle governs single-person decision making but not strategic decision making (i.e., games), remains something of a received view on the possibility of negative-valued information. Over the last half-century decision and game theorists have become keenly aware of the crucial role that act-state independence plays in standard decision theory, aided by a slew of puzzles and aberrant behavior in examples that are found upon close inspection to depend on violations of this independence condition (Kadane, Seidenfeld, and Schervish 2008). Act-state independence is the first thing to go in the theory of games, however, as the whole point of strategic decision making is to factor in the consequences to you from the rational acts of others. So, one might conclude, to avoid the specter of negative-valued information, restrict the scope of Good's principle to single-person decision problems. That, in a nutshell, is the received view on Good's principle. The received view is wrong, however.

2   Good by Savage

To see why single-person decision-making is not immune to negative-valued information, let us consider more carefully Savage's argument that it is immune. To be clear at the outset, our analysis will not uncover a mathematical mistake or faulty theorem. Rather, our aim is to draw attention to another important type of qualification to Good's principle.

Good's principle appears in Savage's discussion in *Foundations of Statistics* of the differences between a basic decision problem and a derived decision problem. A *basic* decision problem is one in which an agent is to choose a basic action from among a collection he judges to be available for choice. A *derived* decision problem is one in which the agent is to choose from the same collection of basic actions but only after considering the associated conditional expected utilities for a basic action given each possible outcome of some experiment. Given the assumption that you wish to maximize your expected utility, why should you prefer a derived decision problem over a basic decision problem? Because you cannot be made worse off in expectation and may well come out better. "It is almost obvious," Savage remarks,

> that the value of a derived problem cannot be less, and typically is greater, than the value of the basic problem from which it is derived. After all, any basic act is among the derived acts, so that any expected utility that can be attained

by deciding on a basic act can be attained by deciding on the same basic act considered as a derived act. In short, the person is free to ignore the observation. That obvious fact is the theory's expression of the commonplace that knowledge is not disadvantageous (Savage 1972, p. 107).

Good later showed that Carnap's principle of total evidence (Carnap 1947) follows as a consequence of the principle to maximize expected utility, so long as the costs of acquiring information are negligible.

[I]n expectation, it pays to take into account further evidence, provided that the cost of collecting and using this evidence, although positive, can be ignored. In particular, we should use all the evidence *already* available, provided that the cost of doing so is negligible. With this provisio then, the principle of total evidence follows from the principle of rationality (Good 1967, p. 319).

Our discussion in the next two sections will be helped along by introducing a bit of formalism now to set up Savage's version of Good's principle.

Following (Pedersen and Wheeler 2015), consider an illustration of Good's principle in Figure 2. Suppose that at some time t_1 you are to face a choice, A, among two courses of action, a_1 or a_2. Prior to this choice you face a decision, O, at some time t_0 prior to t_1, between o_1, the *basic decision* of choosing a_1 or a_2 at time t_1, and o_2, the *derived decision* of choosing a_1 or a_2 at some later time t_2 after you have observed, at no cost, the outcome of an experiment $\mathcal{E}$, with outcomes e_1 or e_2.[1]
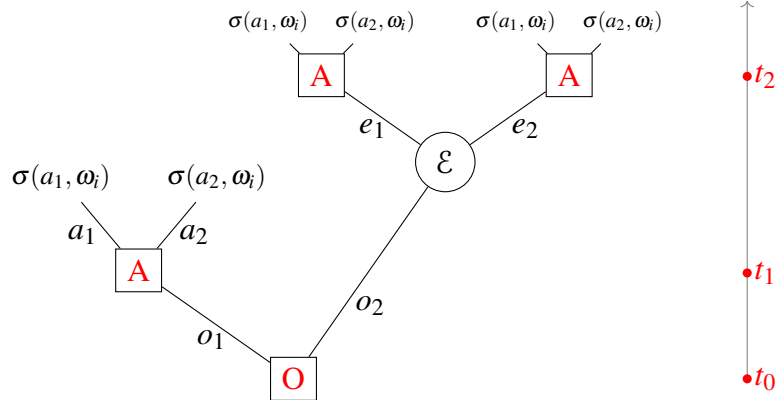

Figure 2: Illustration of Good's Principle

Choice, being governed here by dominance reasoning, comes after ruling out those options for choice that are worse than all others. Those acts that survive the cull are *admissible* for choice. Suppose that your judgments of admissibility can be represented in terms of subjective expected utility maximization with respect to a real-valued expectation $\mathbb{E}_p[\cdot]$ agreeing with a real-valued probability function p defined on a Boolean algebra $\mathscr{A}$ over the set of

---

[1]Good's principle is a synchronic rationality principle, governing here the synchronic choice at t_0 between options o_1 and o_2. Our informal discussion of choices taken at future times ought to be viewed as all hypothetical choices entertained at t_0. Put differently, in choosing between o_1 and o_2, we are comparing at t_0 the consequences from engaging in two lines of suppositional reasoning.

states Ω and a real-valued utility function u defined over the set of consequences.[2] Then, at time t_0 you confront a decision problem $O = \{o_1, o_2\}$. If you implement option o_1 at time t_0, then at time t_1 you will face a decision problem $A = \{a_1,, a_2\}$ without observing the outcome of experiment $\mathcal{E}$. If you implement option o_2 at time t_0, then at time t_2 you will face the same decision problem A after observing the outcome of experiment $\mathcal{E}$.

Abusing notation, let 'o_1' also stand for the event of facing the decision problem A after implementing option o_1 at t_0. (Context should make clear which use of 'o_1' we intend.) In a similar manner, let '$o_2, \mathcal{E} = e_i$' stand for the event of facing the decision problem A after implementing o_2 and observing the outcome e_i of experiment $\mathcal{E}$. The *choice set* $\mathbf{c}$ of admissible options from A for choice given each alternative, written $\mathbf{c}(A|o_1)$ and $\mathbf{c}(A|o_2, \mathcal{E} = e_i)$ for options o_1 and o_2, respectively, may be defined by the following equations:

$$\mathbf{c}(A|o_1) = \arg\max_{a \in A} \mathbb{E}_{p(\cdot|o_1)}\Big[(u \circ \sigma)\big(a, p(d\omega|o_1)\big)\Big] \tag{1}$$

$$\mathbf{c}(A|o_2, \mathcal{E} = e_i) = \arg\max_{a \in A} \mathbb{E}_{p(\cdot|o_2, \mathcal{E}=e_i)}\Big[(u \circ \sigma)\big(a, p(d\omega|o_2, \mathcal{E} = e_i)\big)\Big] \tag{2}$$

where $\circ$ denotes functional composition.

Good's principle assumes that at t_0 you are certain, regardless of whether or not you choose to observe the outcome of experiment $\mathcal{E}$, that you will choose an option $a \in A$ that maximizes your expected utility. This assumption is codified in how admissible choices are determined for each option o_1 and o_2 in Equations 1 and 2, respectively. A second assumption is that your preferences over consequences remain unchanged. A third assumption is that your beliefs given hypotheses accord with Bayesian conditionalization. With all of this in place, Good's principle states that your expectation of (i) *your maximum conditional expected utility of choosing from* A *under option* o_1 is less than or equal to (ii) *your maximum conditional expected utility under* o_2 *of choosing from* A *given experiment* $\mathcal{E}$. Your expectation of (ii) is strictly greater than (i) unless there is an action from A that maximizes conditional expected utility from A regardless of the experimental outcome of $\mathcal{E}$. In other words, unless the experiment $\mathcal{E}$ is irrelevant (i.e., probabilistically independent), then $\mathbf{c}(O) = \{o_2\}$.

3 UNCERTAINTY AND IMPRECISION

According to the canonical theory of synchronic decision-making under risk, a perfectly rational person is one whose comparative assessments of a set of consequences satisfies the recommendation to maximize expected utility. What underpins this claim is the assumption that a person's qualitative comparative judgments of those consequences (aka, *preferences*) are structured in a particular way (*satisfy specific axioms*) to admit a mathematical representation in terms of inequalities of mathematical *expectations*, ordered from worst to best on the real number line. This structuring of preference through qualitative axioms to admit a numerical representation is the subject of expected utility theory (Wheeler 2018, §1.1).

Savage's theory tells us how to represent preference in terms of some pair of numerical probability and numerical utility functions, an ingenious extension of prior work that showed how to quantify each piece separately, principally von Neumann and Morgenstern's

---

[2]Often a uniqueness result for probabilities and utilities accompanies the representation result (asserting, for example, that the probability function is unique and that the utility function is unique up to a positive affine transformation).

numerical representation of utility, which presupposes a numerical probability function (von Neumann and Morgenstern 1944); and de Finetti's numerical representation of probability, which presupposes a cardinal utility function (de Finetti 1974). Let's focus here on probability assessments, which for Bayesians are understood as a person's partial beliefs.

Consider what it means for a person to have a partial belief in the proposition, L, expressing that a particular car is a lemon. What does it mean for a person to have a partial of belief of 0.40 that L is true? According to the Ramsey-de Finetti conception of partial belief, this means the person is *indifferent* between two sorts of hypothetical transactions. The first hypothetical transaction calls on him to *buy* a contract for €0.40 that pays him €1 if the car is a lemon, whereas the second hypothetical transaction calls on him to *sell* such a contract for the same price. Put differently, the first type requires the person to surrender a sure 40 cents for the promise of 1 Euro on the event of A occurring and risk receiving 0 if L does not occur. The second type of transaction requires the person to accept payment of the sure reward of 40 cents in exchange for agreeing to risk paying back 1 Euro on the event of L occurring and paying out nothing—in terms of this contract—otherwise. The choice of price is up to the person, the utility of Euros is assumed to be linear, and the stakes are presumed to be small enough to not bankrupt the person yet large enough for him to care. (De Finetti was a thoroughly pragmatic fellow, a point sometimes lost on his critics.) The price of 40 cents is fair to this person just in case he is indifferent between buying and selling contracts on L at 40 cents. A person is rational just in case there is no possible way to put together a finite set of buy and sell positions on that person's announced fair prices to cause him a sure loss, a return to that person of a value less than zero no matter how the uncertain events in those contracts are resolved.

We rehearse this canonical account in order to introduce a slight generalization. Airport currency exchange counters post different prices for buying and selling trades between a pair of currencies. While they do so primarily to turn a profit, the same idea can be used to express your uncertainty about the event, or events, controlling the payoffs in the contract. So, rather than require decision-makers to post the *same* number for buying as for selling a contract, we wish to allow for the possibility that they post different numbers. Put differently, rather than oblige an agent to give a single two-sided probability for betting on and against the event L, written $P(L)$, we instead oblige the agent to give two one-sided numbers: (i) a one-sided *lower probability* denoting the maximum buying price for a bet on L, written $\underline{P}(L)$; and (ii) a one-sided *upper probability* denoting the minimum selling price for a bet on L, written $\overline{P}(L)$.

Notice that for someone whose fair price is $P(L)$, he will judge any price $\alpha < P(L)$ to buy a bet on L (to bet on L) as desirable. Similarly, prices to sell a bet on L (to bet against L) that are strictly greater than $P(L)$ will likewise be judged desirable. It is only the fair price, the single numerical value of $P(L)$, that marks the agent's indifference. Similar reasoning applies to the one-sided lower probability $\underline{P}(L)$. Any price $\alpha < \underline{P}(L)$ will be judged a desirable prices to bet on L, and any $\alpha > \overline{P}(L)$ will be judged a desirable price to bet against L. The difference is that there are (possibly) two price points where the agent expresses indifference between a sure award and risky reward in the same currency, namely when the buying price for a bet on L is $\underline{P}(L)$ and when the selling price for bets on L is $\overline{P}(L)$. Only when they are the same value is the agent committed to a fair price. Since $0 \leq \underline{P}(L) \leq \overline{P}(L) \leq 1$, one consequence is that any price p offered between the agent's lower and upper probabilities for L, any α such that $\underline{P}(L) < \alpha < \overline{P}(L)$, the agent is neither obligated to sell nor to buy contracts on L.

It is a commonplace to distinguish between risk and uncertainty, an idea that both Knight and Keynes forwarded a century ago (Knight 1921; Keynes 1921). The notion that it is sometimes sensible to permit a bounded range of probability values rather than to insist on numerically determinate probability values is an even older idea, dating back at least to (Bernoulli 1713) and (Boole 1854). But the rich mathematical and philosophical consequences from working out these ideas have only begun to come into focus more recently (Walley 1991; Augustin, Coolen, de Cooman, and Troffaes 2014; Troffaes and de Cooman 2014).

The lower probability model presented above is very basic and supplied with a behavioral interpretation that is very close to the original, canonical model: instead of one number to describe two attitudes, we allow each attitude to have its own number. This slight change, however, from a fair-price model to a buying and selling price model of belief, is enough to put another dent in Good's principle. We turn to see how, next.

4 DILATING PROBABILITIES

What does Knightian uncertainty look like in our bare-bones lower probability model? The short answer is that we have the means to distinguish between indifference and incomparability, and to do so behaviorally in the same simple terms of the canonical Bayesian model. For a longer answer and a consequence, an example.

Suppose there is a ticket that pays to its owner 100 euros on the event of G, *Germany wins the next world cup*. If you owned such a ticket, how much would you demand to part with it? 100 euros would make you whole, so you should at least be indifferent between receiving a sure 100 for the promise of 100 on the event of G being true.[3] Similarly, if you are sure they would lose, the ticket would be worthless to you, so you would find any (positive) price a desirable selling price. Conversely, how much would you pay to buy such a ticket? Here again if you were maximally uncertain (but otherwise abided by the setup for the model), you might not be willing to pay anything for such a ticket. In such a case your lower probability would be 0. If instead you were certain they will win, you would find any price less than 100 euros desirable and be indifferent to owning the ticket and having a 100 euro note in your pocket: for you, being certain of the outcome G, those two rewards are equivalent.

For my part, I would not know how to give a fair price for G. This does not rule out being bullied by a Bayesian into announcing one, but then again that would be a different decision problem. Hypothetically, I would pay up to 10 euros for a chance to win 100 if Germany won the next world cup. They've done it before, I reckon, so there is some chance they could do it next time. On the other hand, if I had such a ticket, what price would I accept to relinquish my chance at 100 euros if they win? Here I might accept nothing less than 90 euros. So, any price between 10 and 90 euros I would neither buy nor sell a 100 euro contract on G. These prices don't have to be symmetric, nor need they be calibrated to a statistical model. This is still a subjective probability model and these are my attitudes toward buying and selling hypothetical 100 euro contracts on G.

Let us introduce some notation to reason with attitudes like the one I have toward G. A *lower probability space* is a quadruple $(\Omega, \mathscr{A}, \mathbb{P}, \underline{P})$ such that Ω is a set of states, $\mathscr{A}$ is an algebra over Ω, $\mathbb{P}$ is a nonempty set of probability functions on $\mathscr{A}$, and $\underline{P}$ is a *lower*

---

[3]Assume a euro today is worth the same euro in the future, or that values are so-adjusted.

probability function on $\mathscr{A}$ with respect to $\mathbb{P}$—that is, $\underline{P}(F) = \inf\{p(F) : p \in \mathbb{P}\}$ for each $F \in \mathscr{A}$. The value $\underline{P}(F)$ is called the *lower probability* of F. The *upper probability function* $\overline{P}$ is then defined by stipulating that $\overline{P}(F) = 1 - \underline{P}(F^c)$ for each $F \in \mathscr{A}$; the value $\overline{P}(F)$ is called the *upper probability* of F. If $\underline{P}(H) > 0$, then conditional lower and upper probabilities are defined as $\underline{P}(E \mid H) = \inf\{p(F \mid H) : p \in \mathbb{P}\}$ and $\overline{P}(F \mid H) = \sup\{p(F \mid H) : p \in \mathbb{P}\}$, respectively.

Now return to the highly uncertain event, G, that Germany wins the next world cup. The upper probability of G is close to 1, $\overline{P}(G) = .9$, and its lower probability is close to 0, $\underline{P}(G) = .1$, such that

$$\overline{P}(G) - \underline{P}(G) = 0.8. \tag{3}$$

Next, imagine a fair coin toss, whose outcomes are heads (H) and tails (H^c). The outcomes of this normal coin flip form a partition, $\mathcal{E} = \{H, H^c\}$, and the same is true of this future championship title, $\mathcal{G} = \{G, G^c\}$. With these preliminaries in place, we rehearse an example from (Seidenfeld 1994) in which a probability estimate of an event becomes less precise upon receiving information about how the tossed coin lands, regardless of whether it lands heads or lands tails.

Since we judge the coin flip to be fair, our expectation of the coin landing heads is the same as our expectation of it landing tails.

$$\underline{P}(H) = \overline{P}(H) = \frac{1}{2} = \underline{P}(H^c) = \overline{P}(H^c). \tag{4}$$

Equation 4 is what a fair price looks like in a lower probability model. We also assume that the outcome of this coin toss landing heads is *independent* of Germany winning this future championship. If any pair of events are probabilistically independent, surely the events *heads* and *Germany wins!* are. So, for each $p \in \mathbb{P}$, we have

$$p(G \cap H) = p(G)p(H) = \frac{p(G)}{2}. \tag{5}$$

Lastly, let F be the event of either G and H both occurring or both failing to occur, namely $F := (G \cap H) \cup (G^c \cap H^c)$. Given our setup, it follows that the probability of F is determinate: that is, $p(F) = \frac{1}{2}$, for all $p \in \mathbb{P}$.

Proof. For each $p \in \mathbb{P}$, observe that

$$\begin{aligned}
p(F) &= p(G, H) + p(G^c, H^c) \\
&= \frac{p(G)}{2} + \frac{1 - p(G)}{2} \qquad \text{[by (5)]} \\
&= \frac{p(G) + 1 - p(G)}{2} \\
&= \frac{1}{2}.
\end{aligned}$$

$\square$

Figure 3 may help to fix intuitions as to why $p(F) = \frac{1}{2}$, for all $p \in \mathbb{P}$, is so by visualizing three probability mass functions that differ with respect to the probability that G. Note that the counter-diagonal is the complement of F, F^c, which is the event that Germany wins if and only if the coin lands tails. Put differently, if Germany winning and the coin landing

heads are each coded as "success" and Germany losing and tails are coded as "failure", the event F says that the coin and Germany both succeed or both fail, whereas the complement event F^c that exactly one of the two succeeds.
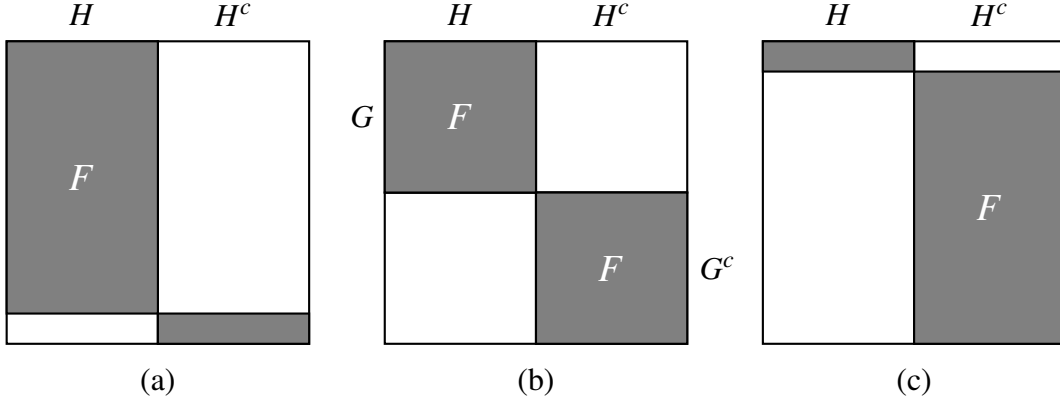

Figure 3: Tables for an uncertain event (row) $\mathcal{G} = \{G, G^c\}$, a fair coin randomizer (column) $\mathcal{E} = \{H, H^c\}$, and the pivotal event (diagonal) F denoting G if and only if H. Figure (a) illustrates when $p(G) = 9/10$ and $p(G^c) = 1/10$, (b) when $p(G) = p(G^c) = 1/2$, and (c) when $p(G) = 1/10$ and $p(G^c) = 9/10$, for $p \in \mathbb{P}$. For each (a), (b), and (c), $p(F) = 1/2$.

Let $\mathcal{E}$ be a positive measurable partition of Ω. We say that $\mathcal{E}$ *dilates F* just in case for each $e \in \mathcal{E}$:

$$\underline{P}(F \mid \mathcal{E} = e) \; < \; \underline{P}(F) \; \leq \; \overline{P}(F) \; < \; \overline{P}(F \mid \mathcal{E} = e).$$

In other words, $\mathcal{E}$ dilates F just in case the closed interval $\left[\underline{P}(F), \overline{P}(F)\right]$ is contained in the open interval $\left(\underline{P}(E \mid \mathcal{E} = e), \overline{P}(E \mid \mathcal{E} = 3)\right)$ for each $e \in \mathcal{E}$ (Walley 1991; Seidenfeld and Wasserman 1993; Pedersen and Wheeler 2014). What is remarkable about dilation is the specter of turning a more precise estimate of F into a less precise estimate, no matter what event from the partition occurs.

Observe that in our World Cup example F is dilated by the coin toss $\mathcal{E} = \{H, H^c\}$: although the initial estimate of F is precisely one-half, learning the outcome of the coin toss, whether heads or tails, dilates the probability estimate of F from one-half to $[.1, .9]$.

Proof. We show that $0.1 = \underline{P}(F \mid H) < \underline{P}(F) = 1/2$.

$$
\begin{aligned}
\underline{P}(F \mid H) &= \inf\{\, p(F \mid H) : p \in \mathbb{P} \,\} \\
&= \inf\left\{ \frac{p\left(\left[(G \cap H) \cup (G^c \cap H^c)\right] \cap H\right)}{p(H)} : p \in \mathbb{P} \right\} \\
&= \inf\left\{ \frac{p(G \cap H)}{p(H)} : p \in \mathbb{P} \right\} \\
&= \inf\left\{ \frac{p(G)p(H)}{p(H)} : p \in \mathbb{P} \right\} \\
&= 0.1
\end{aligned}
$$

$\square$

A similar argument establishes $9/10 = \overline{P}(F \mid H) > 1/2$, and the same argument holds if instead the coin lands tails, i.e., $\underline{P}(F \mid H^c) = 1/10$ and $\overline{P}(F \mid H^c) = 9/10$. Thus, F is dilated by the coin toss, $\mathcal{E} = \{H, H^c\}$.

Here again Figure 3 may help fix intuitions about this result. Notice that the observation of the coin landing heads (H) effectively restricts attention to the first column. Since we learn that H has occurred, the possibilities in the second column associated with tails (H^c) are ruled out. But, the probability mass assigned to the event F in the first column varies widely in Figures 3(a), 3(b), and 3(c). Only in Figure 3(b) does the F have the value $1/2$; Figures 3(a) and 3(c) reveal that the range of uncertainty for F given H is precisely the uncertainty for G displayed in Equation 3. The same argument applies if the coin instead landed tails. As these two outcomes exhaust the possible outcomes of the coin toss, being told that the coin was tossed is enough for a Bayesian to dilate his probability assessment of F. For a discussion of the philosophical and mathematical features of dilation, see (Pedersen and Wheeler 2014; Pedersen and Wheeler 2015).

5 GOOD'S PRINCIPLE AND DILATION

Recall the illustration of Good's principle in Figure 2. Following the presentation in (Pedersen and Wheeler 2015) of an example due to (Seidenfeld 1994), suppose that at t_0 you face a decision problem $O = \{o_1, o_2\}$ where, as before, option o_1 is a basic decision problem A in which you are to choose at t_1 between two acts: a_1, which pays you €1 if E occurs and 'pays' you $-$€1 if E^c, i.e., $\sigma(a_1, F) = $ €1 and $\sigma(a_1, F^c) = -$€1;[4] or the act a_2 which 'pays' you a constant $-$€0.50. Assume that your utility is linear in euro amounts with $u(€x) = x$. Figure 4 fills in these details.

In this basic decision problem A, which is the result of implementing option o_1, the subjective expected utility of a_1 is €0 and the subjective expected utility of a_2 is $-$€0.50. So, a_1 is uniquely admissible from A: receiving nothing is better than paying 50 cents.
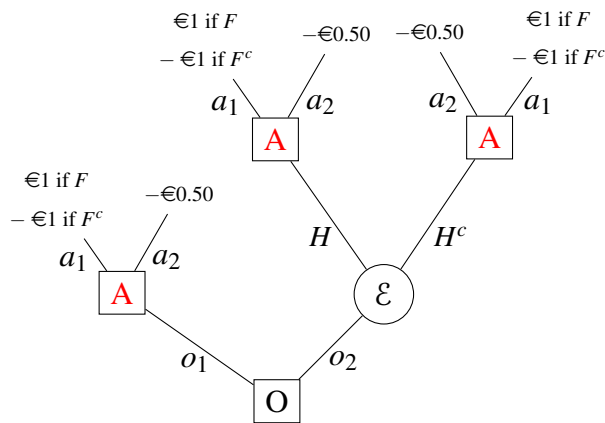

Figure 4: A Sequential Decision Example.

Turn now to option o_2, whereby at t_2 you face a derived decision problem conditional on the outcome of experiment $\mathcal{E}$. Here you are confronted with the same decision problem A at t_2 after learning (only) that H obtains or H^c obtains at t_1. In the derived decision problem act a_1 is inadmissible against a_2. Why? Because in the basic decision problem $p(F) = 1/2$, but in the derived decision problem F is dilated by $\mathcal{E}$ to 0.1 and 0.9: whether the outcome of the fair coin toss is heads or tails, F conditional on that outcome is highly uncertain. Thus, in

---

[4]Here we abuse our notation by writing $\sigma(a, F) = $ €1, for instance, to express that $\sigma(a, \cdot)$ is a constant €1 on F.

the derived decision problem, there are probability mass functions $p \in \mathbb{P}$ whereby $p(F^c)$ is .9, in which case the minimum expected utility of a_1 is $-€0.80$. So, in the derived decision problem, by Savage's Γ-Maximin decision rule, a_2 has a *higher minimum expected value* ($-€0.50$) *than* a_1 ($-€0.80$) regardless of the outcome of the experiment, $\mathcal{E}$.

Assume that a decision maker is certain that she will not change her preferences, will update her belief state by Generalized Bayesian conditionalization (Walley 1991), and that she will choose to maximize her minimal expected utility. Then, in a pairwise choice between a_1 of the basic decision problem determined by option o_1, which has an expected value of zero, and a_2 of the derived decision problem determined by option o_2, which has an expected value of $-\$0.50$, observing cost-free information at t_1, i.e., learning the outcome of the fair coin toss $\mathcal{E}$, is devalued. Here, under the conditions for Good's principle slightly adapted to a lower probability model, we have a case where the decision maker would strictly prefer *not* to receive cost-free information!

6 CONCLUSION

Let's review. The informal version of Bayesian lore has it that it is irrational to turn down cost-free information, since the worst case—when the information is irrelevant to your decision at hand—will leave you at status quo ante. The first restriction to Good's principle is that it does not apply to strategic decision problems, where strategic considerations may disadvantage a player with more information than her opponent. The problem of adverse selection is the classic example, and we discussed Akerlof's market for lemons example and Osborne's formalization. This limitation is fairly well known, however, which is why Good's principle is usually formulated to govern single-person decisions. We then formalized Savage's version of Good's principle in terms of his distinction between a basic and derived decision problem, and where the role maxmin reasoning plays is clear. But, in what may be less widely known, we appealed to the phenomenon of dilation to argue that there are exceptions to Good's principle even for single-person decision problems. Specifically, if one introduces an upper and lower probability model to accommodate a modest form of "Knightian uncertainty", then a probability assessment can become less precise after learning the outcome of an experiment, no matter how that experiment turns out. Finally, we returned to our discussion of basic and derived decision problems in Savage's framework to show that this dilation example can be plugged into Savage's original formulation of Good's principle to show that, by applying Savage's Γ-Maxmin principle, the decision maker would rationally choose to forgo the offer to receive cost-free information about the coin flip experiment $\mathcal{E}$. Thus, for imprecise probabilities, the "commonplace that knowledge is not disadvantageous" is false, even when the costs of obtaining the information is zero. The upshot is that the scope of Good's principle is far narrower than originally conceived and narrower still than many current decision theorists maintain.

The role that Bayesian methods ought to play in models of bounded rationality remains controversial in some circles, and there are some good reasons. Models of bounded rationality typically focus on procedures, algorithms, or psychological processes involved in reaching a decision, securing a goal, or making a judgment, yet these details are ignored in the canonical model. Another branch of bounded rationality focuses on adaptive behavior, and coherent comparative judgments are not, directly at least, the most obvious way to frame this problem.

But it would be incautious to dismiss all of the tools of statistical decision theory, and

unwise to ignore the developments in the field over the last half-century. It is hoped that a wider awareness *better results with less information* results in decision theory—even under the strict adherence to the highly idealized conditions of those mathematical models—will plant a seed of future progress in psychology, where concrete examples are well known. From studying axiomatic departures from the canonical Bayesian theory, it is hoped that the grip of Bayesian dogma will loosen to expand the range of new, creative possibilities for applying a set of practical and powerful mathematical methods (Wheeler 2018).

7   CODA: BLINDED BY OMNISCIENCE

We end with a short remark on logical omniscience. Most formal models of judgment and decision making entail logical omniscience, the presumption that agents have complete knowledge of all that logically follows from their commitments combined together with any and all set of options that are admissible to them for choice. This is as psychologically unrealistic as it is difficult, technically, to remove from formal models. The problem is especially troublesome to Bayesian decision theory, making it difficult to apply the theory to uncertainty about matters of logic and mathematics. Savage, ever prescient, saw the problem that logical omniscience poses to the subjective theory of probability:

> The analysis should be careful not to prove too much; for some departures from theory are inevitable, and some even laudable. For example, a person required to risk money on a remote digit of π would, in order to comply fully with the theory, have to compute the digit, though this would really be wasteful if the cost of computation were more than the prize involved. For the postulates of the theory imply that you should behave in accordance with the logical implications of all that you know. Is it possible to improve the theory in this respect, making allowances within it for the cost of thinking, or would that entail paradox, as I am inclined to believe but unable to demonstrate? (Savage 1967, excerpted from Savage's prepublished draft. See notes in Seidenfeld et al., 2012) .

REFERENCES

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Augustin, T., F. P. A. Coolen, G. de Cooman, and M. C. M. Troffaes (2014). *Introduction to Imprecise Probabilities*. Chichester, West Sussex: Wiley and Sons.

Bernoulli, J. (1713). *Ars Conjectandi*. Basel: Thurnisius.

Boole, G. (1854). *An Investigation of the Laws of Thought*. New York: Dover.

Carnap, R. (1947). On the application of inductive logic. *Philosophy and Phenomenological Research 8*, 133–148.

de Finetti, B. (1974). *Theory of Probability: A critical introductory treatment*, Volume 1 and 2. Wiley.

Gigerenzer, G. and H. Brighton (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science 1*(1), 107–43.

Good, I. J. (1967). On the principle of total evidence. *The British Journal for the Philosophy of Science 17*(4), 319–321.

Grünwald, P. and J. Y. Halpern (2004). When ignorance is bliss. In J. Y. Halpern (Ed.), *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI '04)*, Arlington, Virginia, pp. 226–234. AUAI Press.

Hill, B. (2013). Dynamic consistency and ambiguity: A reappraisal. Technical Report ECO/SCD-2013-983, HEC Paris, Paris.

Kadane, J. B., T. Seidenfeld, and M. J. Schervish (2008). Is ignorance bliss? *Journal of Philosophy 105*(1), 5–36.

Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin.

Lindley, D. V. (1965). *Introduction to Probability and Statistics*. Cambridge: Cambridge University Press.

Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature 27*(4), 1622–68.

Osborne, M. J. (2003). *An Introduction to Game Theory*. Oxford: Oxford University Press.

Pedersen, A. P. and G. Wheeler (2014). Demystifying dilation. *Erkenntnis 79*(6), 1305–1342.

Pedersen, A. P. and G. Wheeler (2015). Dilation, disintegrations, and delayed decisions. In *Proceedings of the 9th Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, Pescara, Italy, pp. 227–236.

Raiffa, H. and R. Schlaiffer (1961). *Applied Statistical Decision Theory*, Volume 1 of *Studies in Managerial Economics*. Harvard Business School Publications.

Ramsey, F. P. (1931). *The Foundations of Mathematics and Other Essays*, Volume 1. New York: Humanities Press.

Savage, L. J. (1967, April). Difficulties in the theory of personal probability. *Philosophy of Science 34*(4), 311–325.

Savage, L. J. (1972). *Foundations of Statistics* (2nd ed.). New York: Dover.

Seidenfeld, T. (1994). When normal and extensive form decisions differ. In D. Prawitz, B. Skyrms, and D. Westerstahl (Eds.), *Logic, Methodology and Philosophy of Science*. Elsevier Science B. V.

Seidenfeld, T., M. J. Schervish, and J. B. Kadane (2012). What kind of uncertainty is that? Using personal probability for expressing one's thinking about logical and mathematical propositions. *Journal of Philosophy 109*(8-9), 516–533.

Seidenfeld, T. and L. Wasserman (1993). Dilation for sets of probabilities. *The Annals of Statistics 21*, 1139–154.

Siniscalchi, M. (2011). Dynamic choice under ambiguity. *Theoretical Economics 6*, 379–421.

Troffaes, M. C. M. and G. de Cooman (2014). *Lower Previsions*. Chichester, West Sussex: Wiley and Sons.

von Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Wakker, P. (1988). Nonexpected utility as aversion of information. *Journal of Behavioral Decision Making 1*, 169–75.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.

Wheeler, G. (2018). Bounded rationality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018 ed.). Metaphysics Research Lab, Stanford University.