The Strategic Role of the Weight of Evidence in Abduction

Lok C. Chan

1 Moore's Argument Against Expected Utility in the *Principia Ethica*

In chapter 1, we saw that Keynes' epistemology was squarely in the Moore-Russell tradition. Keynes' notion of probability as intuitable logical relation is influenced by Moore moral intuitionism, a position that takes 'good' to be an indefinable but intuitable property of an object.[1] Moore is, in particular, well known for his criticism of utilitarianism for committing what he calls the "naturalistic fallacy"[2] The idea is that since 'good' is a basic unanalyzable property, it would be a mistake to, as utilitarians do, explicate it as another natural property such a pleasure or happiness.

This, of course, flies in the face of any mathematical analysis of the right action, which requires a quantitative analysis of the amount of good generated by an action.

[1] G. E. Moore, *Principia Ethica* (Prometheus Books, 1988), 9–10.
[2] Moore, 11–12.

More important, such an approach often requires the use of probability, since an action that produces some good with a low probability ought not be the same as one that generates the same degree of good but with a much higher probability. This is expressed clearly by Pascal and his fellow Jansenists in the *Port Royal Logic* in the 17th century:

> . . . in order to decide what we ought to do to obtain some good or avoid some harm, it is necessary to consider not only the good or harm in itself, but also the probability that it will or will not occur, and to view geometrically the proportion all these things have when taken together.[3]

Moore's naturalistic fallacy can be seen as an attack on the use of the mathematical analysis in ethics, but he also has an argument against the use of *probability* in ethical reasoning. This deeply concerns Keynes. Moore's argument is offered in the chapter "Ethics in Relation to Conduct." His argument is that when we analyze the the good or utility generated by an action, we often only take into account the probability of the *immediate* utility that would follow from the action, but to take utilitarianism seriously, Moore claims, we must consider the effects the action in the long the run.[4] We have to know, for instance, whatever good our action produces in the immediate future will not be negated by the negative effects it has in the long run. However, Moore argues that, because so much about the future is currently unknown to us, we simply lack the ground to make good inferences about what effects each single action will probably case in the long run. For a somewhat crude example, consider a trolley-type problem in which I have to choose between saving a doctor or a criminal. I may reason that

[3] Antoine Arnauld, *Logic, or, the Art of Thinking: Containing, Besides Common Rules, Several New Observations Appropriate for Forming Judgment* (Cambridge University Press, 1996), 273.

[4] Moore, *Principia Ethica*, 158.

saving the doctor will more probably produce more utility in the future, since she will likely save more lives, but this only reflects the short-term outcomes. Why shouldn't I consider, for instance, the probability of the doctor engaging in malpractice and causing a significant amount of suffering, or the probability that the criminal might turn her life around and become a productive member of the society?

Keynes resists this line of argument. To begin, Keynes takes Moore's argument to have demonstrated at most that we cannot know the long term consequence of our action with certainty, but this sets the bar too high—the whole point of employing probabilistic reasoning in our thinking about the right conduct is to maintain a degree of rationality despite of our ignorance about the future. A probability that is based on very little evidence is still "a genuine probability."[5]

Moore's conclusion, as Keynes points out, does make sense from a Frequentist point of view. Most, if not all, of our actions cannot be understood as a part of a long term frequency, so naturally the use of probability to determine the right conduct has a very little meaning from this perspective—if probability is strictly defined as something that expresses itself only in long term behavior, we cannot speak of the probability of an event with any credibility unless we can study the event in a controlled and repeatable environment.

Keynes' logical interpretation, however, does not require the meaning of a probability to be grounded in an empirically verified frequency. A probability is meaningful as long as we rightly perceive the logical relation between the premises and the conclusion. If we have no reason to think that our action is more likely to cause one long

[5]John Maynard Keynes, *A Treatise on Probability* (Macmillan; Co., Limited., 1921), 354.

term consequence than another, then, on Keynes' interpretation of probability, we should rightly regard these events with indifference.[6]

While Keynes does not accept the conclusion of Moore's argument, he is deeply concerned by the incomplete nature of our knowledge of the remote future. To begin, we saw in chapter 1 that the application of Principle of Indifference has very strict conditions, and leads to contradictions when they are not followed, so unless these conditions obtain, we cannot properly calculate the expected value of our action. Keynes is also skeptical that all utilities can be precisely measured.[7]

Keynes also concerns himself with the relationship between evidence and expected utility, and this brings us back to the issue of the weight of evidence.

2 Evidential Weight and Expected Utility

Keynes notes a puzzling dynamic between evidential weight and expected utility—on one hand, expected utility seems to be entirely independent of the weight of evidence, since utility is only discounted by the magnitude of its probability, without any regard to is weight. On the other hand, however, proponents of expected utility theory often implicitly assumes the importance of the weight of evidence. Bernoulli, for instance, suggests that rationality demands the utilization of all evidence available to us. Keynes reasons that this implies that it's always rational to get more evidence, but then it raises another critical question about whether or not one could ever be rational in

[6]Keynes, 354.
[7]Keynes, 356.

refusing new evidence.[8] If the answer for the former question is positive, and the latter question negative, then we have to conclude that rationality dictates us that we should never stop looking for more evidence.

Keynes does not make the jump from "using all the evidence" to "get all the evidence" clear. Nevertheless, this problem is revisited many years later in an exchange between Ayer and Carnap. In his *Logical Foundation of Probability*, Carnap restates Bernoulli's maxim as "the requirement of total evidence".

> *Requirement of total evidence*: in the application of inductive logic to a given knowledge situation, the total evidence available must be taken as basis for determining the degree of confirmation.[9]

Ayer, in response to Carnap, raises the Keynesian question: should "total evidence" include relevant evidence that I do not yet have in possession?[10] The answer must be "yes", Ayer argues. If finding the truth value of some proposition P could potentially sway the balance of my evidence, then I should definitely acquire it. Thus the principle of total evidence seems to suggest that I am also rationally compelled to consider some evidence I do not yet have.

But Ayer points out that this cannot be the whole picture: taken as a rule of rationality, this means we should never stop acquiring unless we are certain that we have acquired all available evidence. This, however, assumes that we know what evidence is available, but it could often be unrealistic to expect to know how much evidence we *do not* currently have.

[8] Keynes, 84–85.

[9] Rudolf Carnap, *Logical Foundations of Probability* (Chicago]University of Chicago Press, 1950), 211.

[10] A. J. Ayer, *Probability and Evidence* (Macmillan, 1972), 56.

Ayer argues that this is only a symptom of a deeper problem about logical nature of probability and its bearing on the rationality of our evidential practice. The logical interpretation of probability, held by both Carnap and Keynes, takes probability as a logical relation between propositions, so within this picture, inductive rationality is a matter of having the right degrees of belief between premises and conclusion—this is analogous to the idea deductive rationality is perceive correctly whether the conclusion follows necessarily from the premises.

Logic, however, does not care about how much evidence we have; it only cares about the relation between our propositions. This problem affects the subjective interpretation of probability as well. As Leonard Savage points out, probability could reveal to us the incoherence within the web of our belief, but it cannot tell us how to resolve it.

> According to the personalistic view, the role of the mathematical theory of probability is to enable the person using it to detect inconsistencies in his own real or envisaged behavior. It is also understood that, having detected an inconsistency, he will remove it. An inconsistency is typically removable in many different ways, among which the theory gives no guidance for choosing.[11]

We encountered a version of this problem in chapter 1, in which we considered Keynes' definition of the weight of evidence in terms of the conditional relevance. Once again, the weight of evidence seems directly relevant to inductive reasoning, yet it cannot be easily situated in the probabilistic framework. The notion of resiliency,

[11] Leonard J. Savage, *The Foundations of Statistics* (Dover, 1954), 57.

which we discussed in detail in chapter 1, does not seem to do any better. While it captures the expression of the weight of evidence, it does very little in illuminating on how it dedicates the rationality of our decisions. Both Joyce and Skyrms are silent on this.

Our original problem of how the severe uncertainty affects the rationality of our current action has seemingly decomposed into two different problems: the original problem about the uncertainty of the remote future and a problem regarding the rationality of our evidential practice. I content that these two problems are in fact both sides of the same coin: the uncertain nature of the remote future is not something we can change in a substantial way—the real question is how we can rationalize our current action by reasonably projecting stability into the future.

The *utility* of evidence, Keynes suggests, is the key of the solution: he suggests that often getting evidence for a belief low in evidential weight will "probably produce the greatest amount of good", but the situation is opposite when the evidence for the belief is weighty—"there clearly comes a point when it is no longer worth while to spend trouble"[12] Thus, for a hypothesis of interest H, the same evidence E generates different amount of utility relative to the amount of information we already have for H. If an agent has almost no information about H, gathering more information would generate the most utility, but for the same evidence, the demand might to low, because the agent might already have enough information about H, so getting more evidence would yield very little to no utility.

If this is right, the importance of the weight of evidence lies not purely in the

[12]Keynes, *A Treatise on Probability*, 84–85.

amount of evidence, but how much we have relative to how much we *need*. In the *Treatise*, Keynes does switch implicitly these two way of thinking of evidential weight—sometimes refers the weight of evidence a balance “between the absolute amounts of relevant knowledge and of relevant ignorance respectively”[13] In a later chapter of the *Treatise*, he also calls weight “the degree of completeness of the information”[14] These remarks suggest that weight is about how we *do not* know as much as how much we *do* know.

This property of evidential weight was already apparent in our analysis of the concept of resiliency. In terms of Skyrms’ conditional resilience, we saw that the more evidence we have, the more resilient a belief tends to get—in the context of the utility of evidence, this means that a resilient belief is one for which the trouble to get more evidence would not be worthwhile.

It is rather unfortunate that Keynes has not further elaborated on this. The idea that the demand of evidence scales with the amount we have, in addition to the problem with the strict definition discussed in chapter 1, should make it relatively clear to Keynes that the weight of evidence cannot increase whenever relevant evidence has been introduced. On the other hand, the relativized notion of weight, implicit from his other remarks, dovetails nicely with the concepts of resilience discussed in chapter 1.

The puzzle about the utility of evidence, and its bearing on the rationality of the gathering of evidence, has been addressed Ramsey on an unpublished note. Interestingly, however, he has proven essentially the opposite conclusion reached by

[13]Keynes, 78.
[14]Keynes, 357.

Ayer and Keynes: Ramsey shows that we should always look more more evidence, because we can never be worse off from doing it. How can this be?

3 The Value of Evidence in Light of the Weight of Evidence

Ramsey's argument is roughly that, *if* we assume an agent to be a perfect Bayesian and that new information does not cost anything, then she will never be no worse off getting new evidence.[15] In fact, she is guaranteed to be *better* off as long as the new evidence will tell her something new. A perfect Bayesian agent is someone who studiously updates her opinions based on Bayes' rule and then act by choosing the action that maximize her expected utility. Note that this assumes two things: first, for any decision problem she faces, there is always going at least one course of action that maximizes her expected utility, and second, as Skyrms points out, this also implies that the agent knows that she will always *stays* being perfectly Bayesian in the future.

I will make use of an intuitive example rather than reproducing the proof here.[16] Suppose we have three hypotheses about the content of an urn in front of us:

1. H_b: 90 black balls and 10 white balls
2. H_w: 10 white balls and 90 black balls

[15] F. P. Ramsey, "Weight or the Value of Knowledge," *British Journal for the Philosophy of Science* 41, no. 1 (1990): 1–4, also see I. J. Good, "On the Principle of Total Evidence," *British Journal for the Philosophy of Science* 17, no. 4 (1966): 319–21 and Savage, *The Foundations of Statistics*, sec 6.2.

[16] This example is adapted from Isaac Levi, "The Weight of Argument," in *Fundamental Uncertainty: Rationality and Plausible Reasoning*, ed. Silva Marzetti Dall'Aste Brandolini and Roberto Sczzieri (Palgrave MacMilan, 2011), 39–58

3. H_n: 50 white balls and 50 black balls.

We then start by assuming $P(H_b) = P(H_w) = P(H_n) = 1/3$. Suppose we win \$1 by picking the correct hypothesis. Our expected payoff for choosing each hypothesis would be the same at 1/3. Nevertheless, we are allow to sample with replacement as many times as we wish. Should we get more evidence? Yes, according to Ramsey, we should.

To begin, at this point, the probability of getting a black ball is the same as getting a white ball. Let E_b be "a black ball is drawn" and E_w for white balls. So:

$$P(E_b) = P(H_b)P(E_b|H_b) + P(H_w)P(E_b|H_w) + P(H_n)P(E_b|H_n)$$

$$= 1/3(0.9) + 1/3(0.1) + 1/3(0.5) = 0.5$$

And $P(E_w) = 1 - P(E_b) = 0.5$. So, in the event of drawing a black ball from the urn, we would update our belief like so:

$$P(H_b|E_b) = \frac{P(H_b)P(E_b|H_b)}{P(E_b)} = \frac{1/3(0.9)}{0.5} = 0.6$$

Similarly, applying the calculation on the other hypotheses, we get:

$$P(H_w|E_b) = 0.067$$

$$P(H_n|E_b) = 0.333$$

Similar argument can be made assuming E_w, that is, a white ball is chosen. In that case $P(H_w|E_w) = 0.6$. If we were an ideal Bayesian agent, we should pick H_b if E_b, and pick H_w if E_w. Since an ideal Bayesian would choose the option that maximizes our expected utility, in either case the expected value after drawing from the urn once is 0.6, which is an improvement, since before drawing our expected utility is 1/3 for all options. The net gain in expected utility would be $0.6 - 0.33 = 0.27$, is referred to as *the value of information* in the decision theory literature.[17]

It turns out that we would be even better off if we were to draw from the urn again. Suppose the first draw yields a black ball. So now we have one piece of evidence in hand. Let us refer to our state of belief after the first draw as $H'_b, E'_b, ..$ and so on. For instance, $P(H'_b) = P(H_b|E_b)$ and $P(E'_b) = P(E'_b|E_b)$. One notable change is that $P(E'_b) = 0.7132$ and $P(E'_w) = 0.2868$. If we draw again and get a black ball, this means:

$$P(H'_b|E'_b) = 0.757$$

$$P(H'_w|E'_b) = 0.009$$

$$P(H'_n|E'_b) = 0.233$$

If a white ball were to be drawn:

$$P(H'_b|E'_w) = 0.21$$

[17] Howard Raiffa and Robert Schlaifer, *Applied Statistical Decision Theory* (Harvard, 1964), 89–90. For a more digestible presentation see Robert Winkler, *An Introduction to Bayesian Inference and Decision* (Probabilistic Publishing, 2010) sec.6.3.

$$P(H'_w|E'_w) = 0.21$$

$$P(H'_n|E'_w) = 0.58$$

Thus, if for the second sample we get a black ball, we would choose b since it has the maximum expected utility at 0.757, and if we get a white ball, we choose n with the expected value at 0.58. So, the expected utility, if we were to draw from the urn again, is: $0.7132(0.757) + 0.2867(0.58) = 0.706$, which is an improvement over just drawing once. The net gain is $0.706 - 0.6 = 0.106$. Ramsey's proof shows that we can keep on getting more evidence and we will never be worse off. In fact, we will be better off as long as there is evidence out there we do not yet have.

What should we make of Ramsey's proof? There are two issues involved here. The first is Keynes' observation that evidence can have a diminishing return, so relevant evidence does not increase weight, and the other is how the weight of evidence bears on the rationality of our action, especially when it comes to the gathering evidence. Ramsey's note provides a good answer for the former, but not the latter.

To begin, Ramsey's contribution here is a way for us to think about the relationship between the *weight* of the evidence in possession and the *value* of the potential new evidence.[18] Ramsey clearly thought is the value of evidence E for hypothesis H as something along the line of the difference between the prior expected utility $EU(H)$ and the posterior $EU(H|E)$.

For instance, we saw that in the example above, the posterior expected utility of

[18] As noted, this is essentially the idea of the value of information in decision theory, but, as a historical note, Ramsey, inspired by Keynes' puzzle about the weight of evidence, has anticipated this development by many years.

the first draw was 0.27 higher than our prior expected utility, and we saw a net gain of 0.106 in expected utility if we were to draw again after drawing a black ball, so we see that the first piece of evidence has a higher value than the second one. What Ramsey's proof demonstrates is that new evidence has a diminishing return—I get a "bigger bang for the buck" for my evidence gathering endeavor when I have less evidence. In light of this, Keynes' example of the balance of the evidence unchanged by the introduction of new evidence is then somewhat incomplete. This explains one of Keynes' puzzle about worthiness of our endeavor to get more evidence in light of the evidence we have in possession.

However, the broader normative question is still unanswered: how should the weight of evidence guide the rationality of our action? To be sure, I do not question that given some assumptions, Ramsey's result will necessarily follow: the same result is proven by both Good and Leonard Savage, so there is no doubt that the result will holds if the assumptions are granted, but that's a big *if*—we have to question if how often these assumptions actually hold.

Ramsey probably understood that information was rarely free. However, Ramsey might have interpreted Keynes' puzzle not as a *decision problem about evidence* but a question regarding its intrinsic value. We saw that Ayer essentially posed the same question to Carnap. History essentially repeated itself when I.J. Good puts forth essentially the answer to Ayer. Interestingly, Good interprets Ayer's as questioning "why... we should bother to make new observations."[19] So, Good seem to think what is needed is a justification for getting new evidence *in general*. Ramsey might

[19]Irving J. Good, *Good Thinking: The Foundations of Probability and Its Applications* (Univ Minnesota Pr, 1983), 178.

have interpret Keynes in the same way. With respect to this version of the problem, the proof makes perfect sense, since it demonstrates that all things being equal we usually end up with better expected utility by considering more evidence. But Keynes' question was about reconciling the general duty to get more evidence and the intuition that evidence gathering for a belief of interest is not always a worthy endeavor.

4 Cost and Urgency

In "Note on the Theory of the Economy of Research", written many years before Ramsey's attempt on the problem, Peirce suggests the crucial problem regarding the relations between evidence and utility is "how with a given expenditure of money, time, and energy, to obtain the most valuable addition to our knowledge"[20]. Thus, as far as Peirce is concerned, the pressing epistemological problem is not to ask whether it is rational to get evidence, but how to get them rationally on the limited resource available to us. I think that this is very much in line with how Keynes has in mind—the problem of the joint satisfaction of the requirement of total evidence and the maximization of expected utility cannot be resolved without considering how cost affects the evidential endeavors. If cost is no object by assumption, the problem is so trivialized that it can hardly be called a problem.

The scenario we imagined in the last section quickly breaks down once we starts to introduce some sort of cost. It was assumed in the example that it costs us neither money nor time to draw from the urn, but suppose it costs us 25 cents for each sample.

[20]Charles S. Peirce, "Note on the Theory of Economy of Research," in *Writings of Charles S. Peirce: A Chronological Edition, Volume 4: 1879–1884* (Indiana University Press, 1989), 72.

This means that we would be gaining only $0.27 - 0.25 = 0.02$ in expected payoff for the first draw, and the second draw would definitely not be worth the additional 25 cents. Or suppose that one dollar is not worthy any endeavor that lasts longer than 15 seconds, and it takes 30 seconds to draw from the urn. As soon as minimally realistic assumptions are introduced, Ramsey's result no longer holds.

Cost might also enter into consideration in different forms, e.g., computational cost or memory. In the same context, Savage ponders over an interesting case that introduces yet another dimension of the problem: consider a very ill person, who is given the option to find out with no cost if the disease she has is mortal. Savage points out that an argument can be made that in this case refusing information could be rational. The thought is that the patient may decide that, based on an assessment of her own personality, she would live the rest of her remaining life in agony if she were to find out that her disease is very serious, whereas she could live relatively happily without knowing. Savage's response is that in this case the information is not really free; it has a *psychological* cost.[21]

Savage's response was intended to paint a counterexample as an explainable exception to the idea that it's always rational to get free evidence, but this sort of implicit cost is the norm, not an exception: at the minimal, any endeavor to seek more evidence will cost us at least time, and the loss of time is the loss of opportunity. Cost might also enter into consideration in different forms, e.g., computational cost or memory. In light of this, Ramsey's demonstration—the value of evidence depends on how much evidence we have and how much evidence there is to get—does not ease Keynes' concern at all, but it makes the question of evidence gathering more pressing:

[21]Savage, *The Foundations of Statistics*, 107.

I might not be getting the most out of my evidential endeavor if it turns out that I could be much better off by examining a different hypothesis, for the very set up for Ramsey's proof requires us to have already chosen *one* hypothesis of interest. But in reality, we often have competing interests. The question is not whether I will get some utility out of gathering evidence for hypotheses in which I have interest, but how I should go about doing it, given the limitation of resource.

Peirce's concept of the weight of evidence is particularly forged to deal with these practical concerns. One definitive feature of Peirce pragmatist philosophy is that any intellectual concept must be clarified and differentiated by their *practical consequences*; otherwise, it should be considered as unfit for the use in any epistemic or scientific endeavor.

This important practical role evidential weight, as I understand Peirce, lies in its function in determining the *urgency* of a hypothesis. To begin, like Keynes, Peirce recognizes that the weight of evidence has much to do with the *amount* of evidence available. In a passage that anticipated in Keynes' distinction between the balance and the weight of evidence, Peirce says:

> to express the proper state of our belief, not one number but two are requisite, the first depending on the inferred probability, the second on the *amount* of knowledge on which that probability is based.[Emphasis added.][22]

However, Peirce takes the evidential weight of a probability judgment to be more conceptually connected to the notion of the *precision* of a probability. Peirce was

[22]Charles S. Peirce, "The Probability of Induction," in *Writings of Charles S. Peirce: A Chronological Edition, Volume 3: 1872–1878* (Indiana University Press, 1966), 295.

impressed by Boole's way of deriving indefinite or imprecise probabilities by using arbitrary constants.[23] He was also interested capture the imprecision of probability using the theory of probable error, which in turn leads him to something very similar to the modern Frequentist notion of confidence interval.[24]

More important, Peirce thinks that the connection between the weight of evidence and precision of probability provides a way in which the utility of evidence can be quantified—the very notion of epistemic progress can be explicated as the decrease in imprecision.[25]

Reference

Arnauld, Antoine. *Logic, or, the Art of Thinking: Containing, Besides Common Rules, Several New Observations Appropriate for Forming Judgment.* Cambridge University Press, 1996.

Ayer, A. J. *Probability and Evidence.* Macmillan, 1972.

Carnap, Rudolf. *Logical Foundations of Probability.* Chicago]University of Chicago Press, 1950.

Good, I. J. "On the Principle of Total Evidence." *British Journal for the Philosophy of Science* 17, no. 4 (1966): 319–21.

[23]Charles S. Peirce, "Harvard Lecture Vi, 1865," in *Writings of Charles S. Peirce: A Chronological Edition, Volume 1: 1857–1866* (Indiana University Press, 1982), 239.

[24]Charles S. Peirce, "A Theory of Probable Inference," in *Writings of Charles S. Peirce: A Chronological Edition, Volume 4: 1879–1884* (Indiana University Press, 1989).

[25]Peirce, "Note on the Theory of Economy of Research," 72.

Good, Irving J. *Good Thinking: The Foundations of Probability and Its Applications*. Univ Minnesota Pr, 1983.

Keynes, John Maynard. *A Treatise on Probability*. Macmillan; Co., Limited., 1921.

Levi, Isaac. "The Weight of Argument." In *Fundamental Uncertainty: Rationality and Plausible Reasoning*, edited by Silva Marzetti Dall'Aste Brandolini and Roberto Sczzieri, 39–58. Palgrave MacMilan, 2011.

Moore, G. E. *Principia Ethica*. Prometheus Books, 1988.

Peirce, Charles S. "A Theory of Probable Inference." In *Writings of Charles S. Peirce: A Chronological Edition, Volume 4: 1879–1884*. Indiana University Press, 1989.

———. "Harvard Lecture Vi, 1865." In *Writings of Charles S. Peirce: A Chronological Edition, Volume 1: 1857–1866*. Indiana University Press, 1982.

———. "Note on the Theory of Economy of Research." In *Writings of Charles S. Peirce: A Chronological Edition, Volume 4: 1879–1884*. Indiana University Press, 1989.

———. "The Probability of Induction." In *Writings of Charles S. Peirce: A Chronological Edition, Volume 3: 1872–1878*. Indiana University Press, 1966.

Raiffa, Howard, and Robert Schlaifer. *Applied Statistical Decision Theory*. Harvard, 1964.

Ramsey, F. P. "Weight or the Value of Knowledge." *British Journal for the*

Philosophy of Science 41, no. 1 (1990): 1–4.

Savage, Leonard J. *The Foundations of Statistics.* Dover, 1954.

Winkler, Robert. *An Introduction to Bayesian Inference and Decision.* Probabilistic Publishing, 2010.