Economics and Philosophy, **35** (2019) 103–158 © Cambridge University Press
doi:10.1017/S0266267118000160 First published online 21 June 2018
cambridge.org/eap

COMPARATIVE VALUE AND THE WEIGHT OF REASONS

ITAI SHER*

Abstract: One view of practical reasoning is that it involves the weighing of reasons. It is not clear, however, how the weights of reasons combine, especially given the logical and substantive relations among different reasons. Nor is it clear how the weighing of reasons relates to decision theoretic maximization of expected value. This paper constructs a formal model of reasons and their weight in order to shed light on these issues. The model informs philosophical debates about reasons, such as the question of whether reasons or values are more fundamental and the opposing doctrines of atomism and holism about reasons.

Keywords: weight of reasons, deliberation, practical reason, decision theory

1. INTRODUCTION

One view of practical reasoning is that it involves the weighing of reasons. When deciding what to do, we have reasons for and against various alternative courses of action, these reasons have weights or strengths, and we choose the action that is favoured by the balance of reasons. It is not clear, however, how the weights of reasons combine, especially given the logical and substantive relations among different reasons. Nor is it clear how this view of practical reasoning relates to the decision theoretic view that practical rationality amounts to the maximization of expected value. This paper constructs a formal model of making a decision on the basis of reasons in order to shed light on these issues. The model informs philosophical debates about reasons, such as the question of whether reasons or values are more fundamental and the opposing doctrines of

* Department of Economics, Crotty Hall, 412 North Pleasant Street, University of Massachusetts, Amherst, MA, 01002, USA. Email: itaisher@gmail.com. URL: https://sites.google.com/site/itaisher/

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Reasons For	Weight	Reasons Against	Weight
R_1	w_1	R_4	w_4
R_2	w_2	R_5	w_5
R_3	w_3	R_6	w_6

TABLE 1. Reasons for and against.

atomism and holism about reasons. There has been some recent work that formally models reasons and other formal work that relates to my model. I will discuss these relations at the end of this introduction.

For most of the remainder of this introduction, I will examine more carefully the view of practical decision-making as the weighing of reasons. I will ask whether this view is coherent in light of the complexity of the relations among reasons. I will schematically outline the sort structure that a theory of weight ought to have. I will postpone discussion of the relation of weighing reasons to maximization of expected value, as in decision theory, until Section 4. This relation will feature prominently thereafter.

So consider practical reasoning as the weighing of reasons. We construct arguments to justify our actions. We weigh reasons for and against certain actions, and decide on the basis of the weight of reasons which action to select.

Let us consider an example to see how this might work. Suppose that I am considering the question: "Should I go to the movies tonight?" I may list the reasons for and against going to the movies along with the weights associated with these reasons as in Table 1.

Among the reasons for going to the movies may be that I promised Jim that I would go (R_1) and that it will be fun (R_2). Among the reasons against may be that I have other work to do (R_4). These reasons have weights. The weight of R_1 is w_1, the weight of R_2 is w_2 and so on. A natural decision rule is:

$$\text{Go to the movies} \Leftrightarrow w_1 + w_2 + w_3 \geq w_4 + w_5 + w_6. \tag{1}$$

In other words, I add up the weights of the reasons for, add up the weights of the reasons against, and go to the movies if and only if the sum of the weights of reasons *for* is greater than the sum of the weights of the reasons *against*.

The idea that weights can be added up in this way is simple and appealing, but such a rule cannot be generally valid. To see this, consider a different example. Suppose that I have a disease. My doctor proposes a treatment, and the question I am considering is, 'Should I take the treatment?' Reinterpret Table 1 as applying to this decision. Suppose that R_1 is now the proposition that the treatment would prolong my life by at

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

least 1 year, and R_2 is the proposition that the treatment would prolong my life by at least 2 years. R_3 through R_6 are other reasons. Can we apply the decision rule (1) to this decision? It seems not. In particular, the sum $w_1 + w_2$ of the weights of reasons R_1 and R_2 does not represent a meaningful quantity. This sum double counts the weight of the fact that the treatment will prolong my life by at least one year, as this fact is entailed by both R_1 and R_2.

Let us consider another example that is yet more complicated. Consider a public health official who must decide whether a treatment will be administered to many people who suffer from a medical condition. R_1 is the proposition that the treatment will prolong the people's lives. R_4 is the conditional proposition that if the treatment prolongs someone's life, that person will suffer harmful side effects. Assuming that R_1 holds, R_4 is a reason against administering the treatment. But imagine instead that R_1 did not hold. Suppose instead that the treatment does not prolong anyone's life; the treatment only reduces the pain caused by the condition. Then R_4 would not really be a reason against the treatment. That R_4 is a reason against administering the treatment depends on R_1 being true. It also follows that the value of the weight w_4 depends on whether R_1 holds. More generally, we can see that the valence of a reason – whether it is a reason *for* or *against* a certain action – can depend on what other reasons obtain.

Let us consider one final example. We return to the question of whether I should go to the movies tonight. As above, R_1 is the proposition that I promised Jim I would go. However, R_2 is now the proposition that it is important to keep one's promises. Plausibly, the fact that it is important to keep one's promises is a reason for me to go to the movies, given that I have promised Jim I would go. It is clear that the weights of R_1 and R_2 cannot be assigned independently. That I promised Jim I would go only has weight if it is important to keep one's promises, and the fact that it is important to keep one's promises is a reason to go only given that I have promised to go. How can I separately assign the weights w_1 and w_2? And even if I could determine such weights, it seems that the sum $w_1 + w_2$ would not represent anything meaningful.

The above examples show that we cannot simply add the weights of reasons, and different reasons are interdependent. Should we regard this as a puzzle? Should we regard it as showing that the notion of attaching weights to reasons does not make sense and that the picture according to which to make a decision, we add the weights of reasons is meaningless? I will argue that the answers to these questions is no.

Consider the question: How do the *weights* of

R_1: The treatment would prolong my life by at least 1 year, and
R_2: The treatment would prolong my life by at least 2 years

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

relate? How does the *weight* of R_2 change once we have already taken the fact that R_1 holds into account? This may seem a bit mysterious.

To see that it is not mysterious, let us consider a different but related question: How do the *probabilities* of

R_1: The treatment would prolong my life by at least 1 year, and
R_2: The treatment would prolong my life by at least 2 years

relate? How does the *probability* of R_2 change once we have already taken the fact that R_1 holds into account? The relation between the probabilities of the propositions R_1 and R_2 may be complicated, but there is nothing specifically mysterious about the above questions. Probability theory deals easily with such questions.

Reflecting on probability theory can provide a useful starting point for developing a theory of weight. In so doing, it is useful to consider three features of probability theory.

1. **Interrelated Propositions**. Probabilities can be assigned to logically and substantively interrelated propositions. However, such probability assignments do not lead to problems of double counting. Sometimes, we can add the probabilities of different propositions.[1] When we cannot add, probability theory tells us how to correct for the overlap.[2] The interrelations among propositions to which we assign probabilities can be complex, but this complexity does not create any in principle barrier to assigning probabilities.
2. **Conditioning and Independence.** The notions of *independence* and *conditional probability* are important for understanding how probabilities are constructed and how different probabilities are related.
3. **Axioms.** There are certain principles or axioms relating probabilities of interrelated propositions. These axioms help to organize our thinking about probabilities.

Each of these features has an analogue for weights: (1) We can assign weights to complex collections of interrelated propositions without double counting. (2) We can define a notion of *independence for weights*

[1] We can add the probabilities of propositions when the propositions are pairwise incompatible.

[2] Let $P(A)$ be the probability of A. For two propositions (modelled as events) A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Subtracting $P(A \cap B)$ cancels out the overlap in the events, and thus avoids double counting. More generally for a collection of events $\{A_1, \ldots, A_n\}$, we use the inclusion-exclusion formula to avoid double counting: $P(\bigcup_{i=1}^{n} A_n) = \sum_{i=1}^{n} (-1)^{i-1} \sum_{S \subseteq \{1,\ldots,n\}, |S|=i} P(\bigcap_{j \in S} A_j)$. Analogous facts hold for calculating the probability of the intersection rather than the union of events.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

and of *conditional weight*. (3) There are axioms that govern the relations of weights of interrelated propositions.

While there are similarities between probabilities and weights, there are also differences. First and foremost, the concept of weight differs from that of probability. We should not expect the same axioms to apply to weight and probability. Another important difference is that we will need to assign weight to normative propositions such as the proposition that it is important to keep one's promises. In contrast, it is difficult to conceptualize assigning probabilities to such propositions.

To summarize: in general, we cannot simply add the weights of reasons. Is this a puzzle, or is it philosophically problematic in any way? No. But we have to find the right picture for thinking about how different reasons and their weights are interrelated. We have to develop the theory of weight in a way that is analogous to the theory of probability. That is what this paper aims to do.

The outline of the paper is as follows. Section 2 asks the question: 'What are reasons?' Section 3 provides a model of reasons. Section 4 shows that choosing the action backed by the greatest weight of reasons has the same formal structure as that posited in decision theoretic models that involve choosing the action that maximizes expected value given one's information. This provides a 'translation manual' between a conception of decision making as optimization of value and a conception of decision making as responding to reasons. An important concept that I introduce for modelling the weight of reasons is *deliberative probability*. Section 5 explicates the idea of deliberative probability. Section 6 answers the question: 'When can the weights of different reasons be added?' To answer this question, the distinction between independent and non-independent reasons is presented. The weights of independent reasons can be added, but not the weights of non-independent reasons. I also define the notion of a *conditional weight*, and show that appropriate conditional weights can always be added. An important question is whether weights or values should be taken as fundamental. Section 3.2 defines weights in terms of values (and deliberative probabilities). Section 7 shows that weights can be taken as fundamental, and values derived. Thus the theory of weight presented here is consistent with taking either weights or values as fundamental and the other as derived. Section 8 concludes, focusing in particular on the bearing of this paper on the theses of atomism and holism about reasons and related issues. An Appendix contains proofs of propositions and discussion of technical issues.

There has been surprisingly little work formally modelling reasons and their weight.[3] Horty (2003, 2007, 2012), Cariani (2013), Brown (2014)

[3] For a recent anthology of philosophical work on reasons and their weight, see Lord and Maguire (2016).

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

and Dietrich and List (2016) are exceptions. Most closely related is Dietrich and List (2013). Dietrich and List's work is important because it connects philosophical work about reasons to decision theory. Prior to their work, this important connection had not been explored. Indeed, it is still a very under-explored area and much work remains to be done. This paper was in part inspired by Dietrich and List's work and contributes to the enterprise they initiated. The current paper differs from Dietrich and List in important ways. First, the appeal to the structure of conditional expectations to account for the weight of overlapping reasons in a way that avoids double counting is new here. Second, while Dietrich and List discuss independent reasons, whose weights can be added, they do not have a formal result analogous to my Theorem 1 characterizing independent reasons. Dietrich and List conceive of reasons as properties of alternatives or as propositions about alternatives. I conceive of reasons as propositions that are not necessarily about the alternatives being chosen. My work also introduces novel conceptual notions such as deliberative probability and conditional weights. The current paper draws on Bolker's 1966 characterization of conditional expectations and Bolker-Jeffrey decision theory (Bolker 1967; Jeffrey 1990). This connection is spelled out in Section 7. There is also a connection to confirmation theory. Both Spohn (2012) and Nair (2016) suggest qualitative and quantitative relations between a proposition A being a reason to believe a proposition B and a proposition A confirming a proposition B. This is related to my formal treatment of a proposition R being a reason for an action a and the weight of that reason. Nair (2016) specifically discusses confirmation in relation to the accrual of the weight reasons and the question of when it is appropriate to add the weights of different reasons. Appendix 2 discusses in more detail some of the relations to the work discussed in this paragraph.

2. WHAT ARE REASONS?

This paper is concerned with the structure of the weights of reasons, and specifically, the interrelations among weights of different reasons. It does not address the question of what reasons are. Nevertheless, it is worthwhile to reflect on what reasons might be before proceeding. If we are to explore the structure of reasons, it is good to have a sense of the nature of reasons.

To be concrete, imagine that I am considering some action such as *going to the movies*, *going to war* or *donating money to the needy*. Possible reasons to take such actions include the fact that *I promised Jim I would go*, the fact that *we have been attacked*, and the fact that *there are people in need*. What does it mean for one of the latter propositions to be a reason for one of the former actions? Opinion is split on this. Raz (1999) writes,

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

'We can think of [reasons for action] as facts, statements of which form the premises of a sound inference to the conclusion that, other things equal, the agent ought to perform the action.' Kearns and Star (2009) write, 'A reason to ϕ is simply evidence that one ought to ϕ'. Broome (2013) writes, 'A pro toto reason for N to F is an explanation of why N ought to F'.[4] Alvarez (2016) writes, 'But what does it mean to say that a reason 'favours' an action? One way of understanding this claim is in terms of justification: a reason justifies or makes it right for someone to act in a certain way.'

By the above accounts, normative reasons may be conceived of as *premises* in an argument for an action, as *evidence* that an action ought to be taken, as *explanations* of the fact that an action ought to be taken, or as *justifications* for the choice of an action.[5] Because the goal of this paper is to present a formal framework for modelling reasons that can be useful for theorists with a broad range of views, I will not take a strong stance about the proper interpretation of reasons here. I believe that my framework is compatible with each of these interpretations, or at least that *prima facie* it is not obviously incompatible with any of them. The justificatory interpretation is most in line with the way I speak of reasons below. The evidential interpretation fits particularly well with the formalism. However, because the evidential interpretation is philosophically contentious, and I have my own reservations about it, I do not wish to commit to it. Once the formal account of the structure of reasons is developed, we can return to the question of how well it fits with each of the interpretations of reasons mentioned here. That task is left to future work.

3. MODELLING REASONS

3.1. A First Pass

I begin by presenting a schematic first pass of my model of weight. This schematic outline will be fleshed out more formally in Section 3.2. One reason for splitting the presentation of the model into two steps is that I want to emphasize that there are two separate proposals here: (i) a schematic proposal for a theory of weight and (ii) a particular way of filling out the schematic proposal. There may be other ways to fill out the schematic proposal.[6] A second reason is that the detailed approach presented in Section 3.2 depends on a specific way of modelling value,

[4] In this paper I am concerned with *pro tanto* reasons. Broome (2013) also takes *pro tanto* reasons to be parts of a certain kind of explanation – namely, a weighing explanation. I cite Boome's account of *pro toto* reasons because it is simpler.

[5] There are also reasons for belief and reasons for other attitudes. In this paper, I focus on reasons for action.

[6] I will argue in Section 3.2 that it is important to fill out the schematic proposal in *some* way, but there may be more than one way in which it can be filled out.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

but I want to show that the basic idea does not inherently appeal to this conceptualization of value.[7] Now on to the schematic proposal.

Suppose that I am contemplating the decision between a default action d and an alternative a. Let $R_1, \ldots, R_n$ be a set of propositions, each of which is a potential reason for selecting either the alternative or the default.

In this paper, I will attempt to understand reasons and their weight by considering the effect that coming to know that those reasons obtain have on my decision. This means that I move from a state of uncertainty to a state of knowledge.

Let us consider two types of uncertainty.

Actual Uncertainty. I don't know whether any of the potential reasons $R_1, R_2, \ldots, R_n$ obtain.

Deliberative Uncertainty. In fact, I do know that $R_1, R_2, \ldots, R_n$ obtain, but I consider my decision from a hypothetical position in which I do not know whether these reasons obtain.

I will focus primarily on deliberative uncertainty. The reason is that in making decisions we are often already aware of various considerations that bear on our decision, and our task is to assess their bearing and weight. The approach explored here is to consider the effect of coming to know these considerations, that is, of moving from a state of *hypothetical* ignorance to a state of knowledge. The situation in which we know and want to assess the weight of what we know is more basic than the situation in which we are actually uncertain, and we would like to understand the more basic situation first. I will call the situation in which an agent pretends to be ignorant of the reasons that she possesses for the purpose of assessing the weights of her reasons the *deliberative position*.

Suppose that in the deliberative position, I view the default d and the alternative a as equally good: I have no more reason to choose one than the other. So I take an attitude of indifference.

If I were to learn only that proposition R_i held, this might break the indifference. Perhaps R_i will shift the decision in favour of a. Let us imagine that we could quantify *how much better* I should take a to be than d upon learning R_i. Let $w_a(R_i)$ be a real number that represents how much better I should take it to be to select a rather than d *overall* were I to learn R_i (and *only* R_i).

[7] The explanation of weights in this section does appeal to some notion of value insofar as it appeals to the degree to which a should be taken to be *better* than b. However, the notion of weight could be taken to be basic and the 'betterness' notion could be defined in terms of weight.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Under the *evidential interpretation* of reasons, we might think of R_i as evidence that a is better than d. Under a *justificatory interpretation* of reasons, we might think of R_i as a fact that makes a better than d in a certain respect. Thinking about how learning that R_i obtains would affect what I should do helps me to get clear on the force of this justification. Alternatively, the force of R_i on the decision might consist in how I should rationally respond to learning R_i in the deliberative position.

If R_i is a reason for a over d, then R_i will shift the decision toward a, and this is represented by the condition that $w_a(R_i) > 0$. If R_i is a reason against a, then R_i will shift the decision away from a, and $w_a(R_i) < 0$. The absolute value $|w_a(R_i)|$ represents the *magnitude* of the reason, that is, how much better or worse a should be taken to be than d; the sign of $w_a(R_i)$ – that is whether $w_a(R_i)$ is positive or negative – represents the *valence* of the reason, that is whether the reason makes a better or worse in comparison to d.

Imagine that we start with the function w_a, which assigns to each potential reason R_i a number $w_a(R_i)$. Then we may *define* R_i to be a *reason for* a over d if $w_a(R_i) > 0$, and R_i to be *reason against* a over d if $w_a(R_i) < 0$. We may define $w_a(R_i)$ to be the *weight* of the reason R_i for or against a.

Observe that the notion of weight that we have adopted here is essentially *comparative*. It reflects how R_i shifts the comparison between a and d, not how R_i changes the evaluation of a in an absolute sense. The reason for taking a comparative approach will be explained in Section 3.2.[8]

Let $\bar{R}_i$ be the negation of R_i. Then $w_a\left(\bar{R}_i\right)$ represents how the comparison between a and d should change were one to learn that R_i does not hold. Intuitively, $w_a(R_i)$ and $w_a\left(\bar{R}_i\right)$ are of opposite sign. So if $w_a(R_i) > 0$, it is intuitive to expect that $w_a\left(\bar{R}_i\right) < 0$: If R_i is a reason *for* choosing a over d, then $\bar{R}_i$ is a reason *against* choosing a over d. We will return to this.

Suppose that in addition to the default d, there are multiple other actions $A = \{a, b, c, \ldots\}$. Then we can define $w_a(R_i)$, $w_b(R_i)$, $w_c(R_i)$, … as the weight of R_i for/against a over the default, the weight of R_i for/against b over the default, and so on. We can also examine comparisons of pairs of actions not including the default. For example, $w_{a,b}(R_i)$ may be the weight of R_i for/against a over b.

3.2. A More Detailed Approach to Modelling Reasons

The previous section presented a schematic model of the weight of reasons. The schematic model was presented in the service of the following goal.

[8] I said in the opening paragraph that the basic proposal of this section can be separated from the detailed filling out of Section 3.2. In line with this, the basic reason given in Section 3.2 for making weight comparative can be articulated without the detailed machinery presented there.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Goal: To develop a theory of weight on a par with probability theory that details the structure of weight, and specifically the relations among weights of interrelated propositions.

To achieve this goal, we must put more structure on the weights of reasons than we established in Section 3.1. But what additional structure are we to impose? Recall from Section 3.1 that $w_a(R_i)$ measures the effect on the comparison between an alternative a and the default d of learning that R_i obtains. A natural approach to our task is to first ask what this learning consists in and how it affects the relative values of the alternative and default, and then to try to use the answer to specify the structure of weight. The simplest and most straightforward way of modelling learning is as *conditioning* on the reason. Here conditioning is understood as the conditioning of expected value as in probability theory. We start by assigning a value to the alternative and the default. These values are expected, or average, values because they depend on various conditions, and we average the value over the relevant conditions. We learn R_i. Then we condition the values of the alternative and default on R_i. This conditioning affects the comparison between the alternative and default, and the effect is the weight $w_a(R_i)$.

As I mentioned, the motivation for this strategy is that it uses simple and straightforward tools that are well understood. An additional advantage is that this model immediately translates the notions of reason and weight into the language of decision theory, and thus promises to provide a link between deciding on the basis of reasons and decision theoretic models that involve the maximization of expected value. This approach also immediately imposes a rich structure on weight; this is the structure of conditional expectations. It was our goal to derive a rich structure. We can then check whether this additional structure is reasonable.

I now proceed to describe this approach more precisely.

Let V_d be the value of the default and V_a be the value of the alternative. These values depend on the state of the world. Formally, let Ω be the set of states of the world. V_a then maps each state in Ω to a real number. The same applies to V_d. V_a and V_d are *random variables*. (Random variables are just functions whose values depend on the state of the world.) There is a probability measure P over Ω that assigns probabilities to events (subsets) of Ω.[9] Define $\widehat{V}_a$ by $\widehat{V}_a = V_a - V_d$. $\widehat{V}_a$ is the difference between the value of a and the value of d. Since V_a and V_d depend on the state, in general,

[9] I employ the ordinary set theoretic machinery: There is a sigma algebra on Ω, and P assigns a probability to each event in the sigma algebra.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

$\widehat{V}_a$ depends on the state as well. I refer to $\widehat{V}_a$ as the *comparative value* of a, since taking a difference, it measures the value of a relative to the value of d. $\widehat{V}_a$ measures the comparison between a and d.

On a substantive level, I wish to remain neutral about what constitutes value. It may amount to or include human welfare, under the various conceptions of welfare, and it may include other values, such as those of fairness, freedom, rights and merit. It may represent moral or prudential value. Similarly, the 'outcome' that would result from taking action a is interpreted broadly to potentially include information about rights, merit, intentions and so on.

The *expected comparative value* of a is $\mathbb{E}\left[\widehat{V}_a\right]$. Expectation is defined in the ordinary way. For example, if there are three states, ω_1, ω_2 and ω_3, then

$$\mathbb{E}\left[\widehat{V}_a\right] = \widehat{V}_a\left(\omega_1\right) P\left(\omega_1\right) + \widehat{V}_a\left(\omega_2\right) P\left(\omega_2\right) + \widehat{V}_a\left(\omega_3\right) P\left(\omega_3\right).$$

Here $P(\omega_i)$ is the probability of state ω_i and $\widehat{V}_a\left(\omega_i\right)$ is the comparative value at state ω_i. More generally, $\mathbb{E}\left[\widehat{V}_a\right] = \int_\Omega \widehat{V}_a \, dP$. So the expected comparative value is just a weighted average of the comparative value of a at different states, where the weights in the weighted average are just the probabilities of the states.

There is a question of how we interpret the probability measure P. We can give P one of the standard interpretations, as either objective or subjective probability. However, I will focus on a special interpretation, which I call *deliberative probability*. Deliberative probability corresponds to the situation of deliberative uncertainty discussed in Section 3.1. Deliberative probability is a hypothetical probability constructed for the purpose of deliberation. I will elaborate on deliberative probability at greater length in Section 5.

A *potential reason* is just a proposition, which is modelled formally as just a set of states (a subset of Ω). If R is a potential reason, I assume that $P(R) > 0$. This assumption is for simplicity. One may want to posit additional restrictions on which propositions can count as potential reasons. For example, if one does not like the evidential account of reasons, one might want to insist that a reason is not merely evidence that favours taking an action, but stands in some stronger relation to the action such as justifying the action. So there may be some initial filtering operation that eliminates some propositions as not being potential reasons. In the interest of simplicity, I ignore this filtering operation here.

However, I do make one assumption about potential reasons: I assume that the truth of all potential reasons is determined before any

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

action is chosen.[10] I also assume that if, in deliberation, the agent adopts an attitude of uncertainty toward a potential reason R, then she does not view her choice of action as influencing the probability of R. Thus, if there are propositions whose truth (or the probability that they obtain) is influenced by the choice of action, these propositions are not potential reasons. For example, the fact that I promised is a potential reason. A (non-material) conditional proposition such as 'if Ann were to choose a, then Bob would be harmed' can be a potential reason (not to choose a) since the probability assigned to this proposition is not influenced by whether Ann chooses a or not.[11] In contrast, the proposition 'Bob is harmed' is not a potential reason if whether Bob is harmed depends on whether Ann chooses a.

As the example of 'Bob is harmed' shows, it is natural to assume that the probabilities of propositions describing the consequences of actions are influenced by the choice of action. I implicitly model the effects of actions on consequences through the different values that different actions have at various states. For example, it may be that $V_a(\omega) \neq V_b(\omega)$ because a and b have different consequences at ω. With a little additional machinery it would be straightforward to explicitly model the influence of actions on the probabilities of consequences, but to minimize notation and space, I choose not to do so.

Having discussed *potential* reasons, I now spell out what it is for a proposition R to be *a reason for* action a. In the deliberative position, I have an estimate of the comparative value of a, $\mathbb{E}\left[\widehat{V}_a\right]$. We may assume, as in Section 3.1, that in the deliberative position, the alternative and default are considered to be equally good. This would be expressed by the condition $\mathbb{E}\left[\widehat{V}_a\right] = 0$: the expected difference between the two actions is zero. I call this assumption *initial neutrality*. It is natural to assume that in the absence of reasons, all actions are equally good. However, this assumption is completely inessential to the analysis. Indeed, sometimes we want consider some specific set of reasons given some fixed background information that favours one action over another so that the alternative and default do not start on even footing.

Now imagine that for some potential reason R, I were to learn R and nothing else. Then I would update my evaluation of action a. The value of action a would become $\mathbb{E}[V_a \mid R]$. If there are finitely (or countably) many

[10] That the truth of a proposition is determined prior to the choice of an action does not mean that it is a proposition *about* something that happened in the past. The proposition, 'it will rain tomorrow' may be determined to be true prior to the choice of action and its truth may be uninfluenced by the choice even though this proposition is about the future.

[11] The proposition 'if Ann were to choose a, then Bob would be harmed' is true in some states in Ω and false in others, and we may associate this proposition with the set of states in which it is true.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

states, $\mathbb{E}[V_a \mid R] = \sum_{\omega \in R} V_a(\omega) \frac{P(\omega)}{P(R)}$. More generally $\mathbb{E}[V_a \mid R] = \frac{\int_R V_a \, dP}{P(R)}$. Similarly, I update my evaluation of the difference between the action and the default, attaining $\mathbb{E}\left[\widehat{V}_a \mid R\right]$.

Definition 1. *Let* R *be a potential reason, and let* a *be an action.* R *is a* reason for a *if*

$$\mathbb{E}\left[\widehat{V}_a \mid R\right] > \mathbb{E}\left[\widehat{V}_a\right]. \tag{2}$$

R *is a* reason against a *if* $\mathbb{E}\left[\widehat{V}_a \mid R\right] < \mathbb{E}\left[\widehat{V}_a\right]$.

It is useful to rewrite (2) more explicitly: R is a reason for a if $\mathbb{E}[V_a - V_d \mid R] \geq \mathbb{E}[V_a - V_d]$. In other words, R is a reason for a if the comparison between a and the default d becomes more favourable to a once R is taken into account.

Definition 2. *The* weight *of* R *for/against* a *is*

$$w_a(R) = \mathbb{E}\left[\widehat{V}_a \mid R\right] - \mathbb{E}\left[\widehat{V}_a\right]. \tag{3}$$

Thus the weight of a reason for/against an action a quantifies the change in the comparison between action a and the default that is brought about by taking R into account. Combining Definitions 1 and 2, we see that $w_a(R) > 0$ if R is a reason for a and $w_a(R) < 0$ if R is a reason against a. This is exactly what we assumed in the schematic model of Section 3.1. So the more detailed model presented here is indeed a way of filling in the schematic model of Section 3.1.

Again, it is useful to unwind the definition.

$$\begin{aligned} w_a(R) &= \mathbb{E}[V_a - V_d \mid R] - \mathbb{E}[V_a - V_d] \\ &= \underbrace{\left(\underbrace{\mathbb{E}[V_a \mid R] - \mathbb{E}[V_d \mid R]}_{\text{alternative vs. default}}\right) - \left(\underbrace{\mathbb{E}[V_a] - \mathbb{E}[V_d]}_{\text{alternative vs. default}}\right)}_{\text{Taking } R \text{ Into Account vs. Not Taking } R \text{ Into Account}} \end{aligned}$$

where the second equality uses the linearity of expectations. This shows that the notion of weight is *doubly* comparative; it measures a difference of differences. It measures the way that taking R into account changes the comparison between action a and the default.

One might wonder why reasons and weight are defined in this comparative way. Why not use non-comparative value V_a rather than comparative value $\widehat{V}_a$? Under this non-comparative approach, R is a

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

reason for a if $\mathbb{E}[V_a \,|\, R] > \mathbb{E}[V_a]$.[12] Imagine that I am deciding whether or not to go on holiday. Suppose that I win the lottery. Winning the lottery improves the value of all options. It makes going on holiday better, and it also makes staying home better. If we used the non-comparative definition, then winning the lottery would be a reason both for going on holiday and for not going on holiday, but that cannot be. What matters is how winning the lottery affects the comparison between going on holiday and not going on holiday. If winning the lottery shifts the comparison between going on holiday and not going towards going, then winning the lottery is a reason for going on holiday.[13]

According to Definition 1, we decide whether R is a reason for a by comparing the (comparative) value of a when R is taken into account to the value of a when R is not taken into account. An alternative approach would be to compare the value of a when the proposition R is true to the value of a when R is false. That is, we may want to say that R is a reason for a if and only if

$$\mathbb{E}\left[\widehat{V}_a \,|\, R\right] \geq \mathbb{E}\left[\widehat{V}_a \,|\, \bar{R}\right]. \tag{4}$$

It turns out that these definitions are equivalent. That is, (4) holds if and only if (2) holds.[14] It turns out, however, that $\mathbb{E}\left[\widehat{V}_a \,|\, R\right] - \mathbb{E}\left[\widehat{V}_a \,|\, \bar{R}\right]$ would not be a good definition of weight.[15]

Definitions 1–2 are definitions of one's *pro tanto* reasons and their weight. A *pro tanto* reason is a (not necessarily decisive) consideration that counts in favuor of or against some action. In contrast a *pro toto* reason is a complete and decisive all things considered reason for or against an action.[16] Let $R_1, \ldots, R_n$ be the reasons bearing on the decision. As reasons are here formally modelled as sets of states, we represent the conjunction

[12] I said above that weight and reasons are *doubly* comparative; if we use non-comparative value, then reasons and weight are still comparative in that in the determination of whether R is a reason for a and of the weight of R, consideration of the reason is still compared to the absence of the reason, but no longer comparative in that the action is no longer compared to the default. So with the replacement of non-comparative value for comparative value, reasons and their weight go from being doubly comparative to being singly comparative.

[13] I am grateful to Anna Mahtani for raising this issue. This led me to formulate the theory in terms of comparative rather than non-comparative value.

[14] This follows from the law of iterated expectations.

[15] In particular, the definition of weight (3) leads to natural notions of conditional weight and independence of weights that support the proposition that the weights of independent reasons can be added, as can the conditional weights of non-independent propositions. See Section 6. The alternative definition would not support these constructions and results. The alternative definition would also violate the *Averaging* axiom – see Section 7 – which is a highly compelling axiom for weight.

[16] The completeness of the reason may be relative to the agent's knowledge.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

of reasons as their intersection $\bigcap_{i=1}^{n} R_i$. If the default and alternative are the only options, and we assume initial neutrality, then if $w_a\left(\bigcap_{i=1}^{n} R_i\right) > 0$, then $\bigcap_{i=1}^{n} R_i$ can be regarded as a *pro toto* reason in favour of a.[17] This interpretation will be supported by the analysis of Section 4.

Imagine that there are many actions available, a, b, c, and so on. Then we can define V_b and V_c, $\widehat{V}_b$ and $\widehat{V}_c$ and $w_b(R)$ and $w_c(R)$ analogously to V_a, $\widehat{V}_a$ and $w_a(R)$. The comparative values $\widehat{V}_a$, $\widehat{V}_b$ and $\widehat{V}_c$ and weights $w_a(R)$, $w_b(R)$ and $w_c(R)$ are all defined relative to the same default. We can also focus on comparisons between non-default alternatives. Define $V_{a,b} = V_a - V_b$ and $w_{a,b}(R) = \mathbb{E}\left[V_{a,b} \mid R\right] - \mathbb{E}\left[V_{a,b}\right]$. We can then say that R is *reason for choosing* a *over* b if $w_{a,b}(R) > 0$ and R is *reason against choosing* a *over* b if $w_{a,b}(R) < 0$. Given these definitions, it follows that[18]

$$w_{a,b}(R) = w_a(R) - w_b(R). \tag{5}$$

While the values $w_a(R)$ and $w_b(R)$ depend on the choice of default, the difference $w_a(R) - w_b(R)$ is independent of which action is chosen as default.[19,20] Along the lines of (5), we have the general relation $w_{a,c}(R) = w_{a,b}(R) + w_{b,c}(R)$.[21] This shows that the model has specific consequences for the relation of the weights of a reason for comparisons between different pairs of actions.[22]

[17] Likewise if $w_a\left(\bigcap_{i=1}^{n} R_i\right) < 0$, then $\bigcap_{i=1}^{n} R_i$ can be regarded as a *pro toto* reason against a.

[18] To see that (5) holds, observe that

$$\begin{aligned}
w_{a,b}(R) &= (\mathbb{E}[V_a \mid R] - \mathbb{E}[V_b \mid R]) - (\mathbb{E}[V_a] - \mathbb{E}[V_b]) \\
&= [(\mathbb{E}[V_a \mid R] - \mathbb{E}[V_b \mid R]) - (\mathbb{E}[V_a] - \mathbb{E}[V_b])] \\
&\quad - [(\mathbb{E}[V_d \mid R] - \mathbb{E}[V_d \mid R]) - (\mathbb{E}[V_d] - \mathbb{E}[V_d])] \\
&= [(\mathbb{E}[V_a \mid R] - \mathbb{E}[V_d \mid R]) - (\mathbb{E}[V_a] - \mathbb{E}[V_d])] \\
&\quad - [(\mathbb{E}[V_b \mid R] - \mathbb{E}[V_d \mid R]) - (\mathbb{E}[V_b] - \mathbb{E}[V_d])] \\
&= w_a(R) - w_b(R),
\end{aligned}$$

where the third equality is derived by rearranging terms.

[19] To see this, suppose that instead of d, we choose some other action $\tilde{d}$ as default. Let $\tilde{w}_a(R)$ be the weight of R for/against a relative to $\tilde{d}$. That is, $\tilde{w}_a(R) = \mathbb{E}\left[V_a - V_{\tilde{d}} \mid R\right] - \mathbb{E}\left[V_a - V_{\tilde{d}}\right]$. $\tilde{w}_b(R)$ is defined similarly. Then in general $\tilde{w}_a(R) \neq w_a(R)$ and $\tilde{w}_b(R) \neq w_b(R)$. However, a derivation identical to that in footnote 18 with $V_{\tilde{d}}$ taking the place of V_d shows that $w_{a,b}(R) = \tilde{w}_a(R) - \tilde{w}_b(R)$.

[20] (5) implies that if we know all weights of the form $w_a(R)$ as a and R vary, we can derive all weights of the form $w_{a,b}(R)$.

[21] The derivation of this relation is similar to the derivation of (5) in footnote 18.

[22] Observe that $w_{b,a}(R) = -w_{a,b}(R)$ so that $w_{a,b}(R) > 0$ if and only if $w_{b,a}(R) < 0$. This is related to the following observation: In many cases, whether it is more natural to say that R is a reason for a or a reason against b depends on the context. In such cases, there seems to be no substantive issue at stake. One way of understanding this in terms of the current

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

4. WEIGHING REASONS AND MAXIMIZING EXPECTED VALUE

We are now in a position to think about the relation between weighing reasons and maximizing expected value conditional on one's information. Suppose that an agent is trying to decide whether to take an action a or an action b. Let $R_1, \ldots, R_n$ be the reasons that are relevant to this decision. Recall that $\bigcap_{i=1}^{n} R_i$ is the conjunction of these reasons. $w_a\left(\bigcap_{i=1}^{n} R_i\right)$ is the weight of the conjunction $\bigcap_{i=1}^{n} R_i$ for or against a. $w_a\left(\bigcap_{i=1}^{n} R_i\right)$ is defined by substituting the conjunction $\bigcap_{i=1}^{n} R_i$ for the single reason R in (3). $w_b\left(\bigcap_{i=1}^{n} R_i\right)$ is defined similarly. Then $w_{a,b}\left(\bigcap_{i=1}^{n} R_i\right)$ is defined via (5) with $\bigcap_{i=1}^{n} R_i$ taking the role of R. Let us redefine *initial neutrality* as the assumption that $\mathbb{E}[V_x] = \mathbb{E}[V_y]$ for all actions x and y (including the default).

Proposition 1. *Assume initial neutrality. Then,*

$$\mathbb{E}\left[V_a \middle| \bigcap_{i=1}^{n} R_i\right] \geq \mathbb{E}\left[V_b \middle| \bigcap_{i=1}^{n} R_i\right] \text{ if and only if } w_{a,b}\left(\bigcap_{i=1}^{n} R_i\right) \geq 0. \qquad (6)$$

In other words, the conjunction of all relevant reasons $\bigcap_{i=1}^{n} R_i$ is a reason for a over b if and only if, conditional on all relevant reasons, the expected value of a is greater than the expected value of b.

On a formal level, there is nothing deep about this result; it follows immediately from definitions. However, on a substantive level, the result is important. It shows that the way that we have set up our model, selecting the action supported by the conjunction of reasons is the same as selecting the action that maximizes expected value conditional on those reasons. Thus the model links conceptions of rational decision-making as optimizing expected value and conceptions of rational decision-making as responding to reasons.

Observe that the inequality on the right-hand side of Proposition 1 features the non-comparative values V_a and V_b of a and b rather than the comparative values $\widehat{V}_a$ and $\widehat{V}_b$. Thus the proposition establishes a relation between taking the action that maximizes expected *non-comparative* value and the action supported by the weight of reasons.[23] That we can appeal to non-comparative value here reflects the fact that the choice of default does not matter for decision-making.[24] Indeed, for any set of actions, the

model is that the more natural locution depends on which action is more naturally viewed as the default. I thank an anonymous referee for this observation.

[23] Note, however, that the proposition would still be true if the comparative values were substituted for the non-comparative values.

[24] Snedegar (2018) has recently argued that if 'reasons against' are defined in terms of 'reasons for', then various ways of specifying what reasons against are reasons for are problematic. In my model, reasons for and reasons against are both defined comparatively relative to a comparison between an alternative and a default and are interdefinable. The

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

action in the set most supported by the combined weight of reasons – see Proposition 2 below – is independent of the choice of default.

Proposition 1 assumed initial neutrality. However, this assumption is unimportant; a very similar proposition can be formulated without initial neutrality. If action b started with a default advantage over a of magnitude $\delta_{b,a}$, then Proposition 1 could be modified to say that the expected value of a is greater than that of b conditional on all reasons if and only if the conjunction of all reasons favours a over b and their combined weight exceeds the default advantage of b over a.[25]

Consider again that there may be many actions $a, b, c, \ldots$ and so on. The default is included among these actions. It is possible to compare any pair of actions a and b on this list. Let us suppose for simplicity that the same set of reasons $R_1, \ldots, R_n$ is relevant to every such comparison. I will return to this assumption below. Now we can define a relation $\succsim$ on the set of actions by[26]

$$a \succsim b \;\Leftrightarrow\; w_{a,b}\left(\bigcap_{i=1}^{n} R_i\right) \geq 0. \tag{7}$$

We can interpret $\succsim$ as saying that the conjunction of relevant reasons decisively favour a over b. The following is an immediate consequence of Proposition 1:

Proposition 2. *Assume initial neutrality. The relation* $\succsim$ *is transitive and complete. So, the model does not allow cycles and if there are finitely many actions, there exists an action* a *such that for all actions* b, a$\succsim$b. *Thus, barring ties, there exists an action* a *such that the conjunction of reasons decisively favour* a *over all other actions.*

comparative aspect is consonant with contrastivism, which Snedegar (2015) explicates. I would argue that defining reasons against comparatively with reference to a specific default has advantages over the options considered by Snedegar (2018). Note that as just explained the choice of default does not matter for decision-making. See also the last paragraph of Section 3.2.

[25] More formally, without initial neutrality, let $\delta_{b,a} = \mathbb{E}[V_b] - \mathbb{E}[V_a]$ be the *default advantage* of b over a. Then (6) becomes:

$$\mathbb{E}\left[V_a \,\middle|\, \bigcap_{i=1}^{n} R_i\right] \geq \mathbb{E}\left[V_b \,\middle|\, \bigcap_{i=1}^{n} R_i\right] \textit{ if and only if } w_{a,b}\left(\bigcap_{i=1}^{n} R_i\right) \geq \delta_{b,a}.$$

[26] The relation $\succsim$ defined by (7) is well defined regardless of whether we assume initial neutrality. However, without initial neutrality, $a \succsim b$ does not mean that the reasons *decisively* support a over b. Rather $a \succsim b$ means that the reasons shift the comparison toward a; however the support may not be decisive because it may not overcome the default advantage $\delta_{b,a}$ of alternative b (see footnote 25). If, in the absence of initial neutrality, we want a relation $\succsim'$ that expresses decisive support, then this relation should not be defined via (7), but rather by $a \succsim' b \;\Leftrightarrow\; w_{a,b}\left(\bigcap_{i=1}^{n} R_i\right) \geq \delta_{b,a}$.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

The proposition says that the weight of reasons orders actions for decision-making.[27] Without initial neutrality, a relation that is analogous to $\succsim$ can be defined, and one can derive an analogue of Proposition 2.[28]

The above proposition depends on the assumption that the same set of reasons $R_1, \ldots, R_n$ are relevant to the comparison between every pair of actions. This may not always be true. Suppose, for example, that the reasons are R_1, R_2, R_3. Suppose R_1 and R_2 are relevant to the comparison a and b; R_2 and R_3 are relevant to the comparison of b and c; and R_1 and R_3 are relevant to the comparison of a to c. In comparing a and b, what is relevant is $w_{a,b}(R_1 \cap R_2)$. However, because R_3 is not relevant to the comparison, presumably taking R_3 into account has no effect on the aggregate weight of reasons, so that $w_{a,b}(R_1 \cap R_2 \cap R_3) = w_{a,b}(R_1 \cap R_2)$. If this equality did not hold, then R_3 would be relevant to the decision after all. So even if different reasons are relevant to different comparisons, we can still use $w_{a,b}\left(\bigcap_{i=1}^{n} R_i\right)$ to represent the aggregate weight of all reasons; adding irrelevant reasons has no effect.

In this section, I have explained that the conception of weighing reasons presented here makes weighing reasons formally equivalent to maximizing expected value conditional on one's information. The order $\succsim$ therefore captures both (conditional) expected value and the aggregate weight of reasons. However, there remains a question about when we can add the weights of different reasons. In other words, when is it the case that $w_{a,b}\left(\bigcap_{i=1}^{n} R_i\right) = \sum_{i=1}^{n} w_{a,b}(R_i)$ so that the recommended action is the one supported by the sum of weights of reasons? That question will be answered in Section 6.

Before discussing the conditions under which weights can be added, I will discuss the notion of deliberative probability, which is essential to the model. This will be done in the next section.

I conclude this section by mentioning a caveat about Proposition 1: This proposition shows that weighing reasons and maximizing expected value conditional on information are equivalent in *form*; however, in substance, the conditional expectation in question is a conditional expectation with respect to deliberative probability, rather than with respect to objective or subjective probability. Actually, the matter is a bit more subtle. Recall that the reasons $R_1, \ldots, R_n$ are assumed to be known. We could assume that only the joint probability over these known reasons $R_1, \ldots, R_n$ is hypothetically constructed deliberative probability, and once

[27] It follows from what has been said above that this order is independent of which action is chosen as the default.

[28] The relation $\succsim$ would still be reflexive and transitive without initial neutrality. Initial neutrality is important only for the interpretation of $\succsim$ as *decisive* support. Without initial neutrality then the relation of decisive support is the relation $\succsim'$ defined in footnote 26. Everything that Proposition 2 claims about $\succsim$ under default neutrality is true of $\succsim'$ without default neutrality.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

these reasons have been taken into account all remaining uncertainty is ordinary (subjective or objective) uncertainty, so that if Q is a proposition that is still uncertain once these reasons have been taken into account, $P\left(Q \left| \bigcap_{i=1}^{n} R_i \right.\right)$ is ordinary probability. This sort of mixed deliberative-ordinary probability could be formalized, but in the interest of space, I refrain from doing so here. The simplest case is the one in which there is no relevant residual uncertainty once $R_1, \ldots, R_n$ have been taken into account. To model this, we could assume that there is only one state $\omega*$ compatible with the conjunction of $R_1, \ldots, R_n$. In this case (6) reduces to: $V_a(\omega*) \geqslant V_b(\omega*)$ if and only if $w_{a,b}\left(\bigcap_{i=1}^{n} R_i\right) \geq 0$. Under this assumption, maximizing expected value reduces to maximizing value.

However the above interpretive issues are resolved, Proposition 1 is significant: It says that choosing the action supported by the balance of reasons is analogous to choosing the action that maximizes expected value conditional on one's information. In the deliberative position the agent constructs hypothetical deliberative probabilities – possibly integrated with ordinary probabilities as described in the preceding paragraph, takes reasons into account by updating her probabilities, and chooses the action that maximizes expected value. The change in expected values caused by updating on her reasons is the weight of those reasons. The expected value maximizing action is the action most supported by her reasons. We can use this expected value model to understand the combined weight of all of the agent's reasons or to examine the weight of any subset of reasons and the action supported by that subset.

5. DELIBERATIVE PROBABILITY

5.1. How Can We Interpret Deliberative Probability?

The construction of weights in Section 3.2 depends on deliberative probability. But how do we assign deliberative probabilities? To get a grip on this operation, imagine assessing the value of a large tract of land. You conduct a test to check whether there is oil beneath the land. The test can come back positive or negative. Suppose that the test comes back positive and this greatly increases your estimate of the value of the land. If the test had come back negative, this would not have changed your estimate very much. This means that you must have initially assigned a *low probability* to there being oil under the land because if you had assigned oil a high probability, then you would already have factored the likelihood of oil into your assessment of the land and it would have been the discovery of no oil that would have moved your assessment. The possibility that there is oil has *high significance* in the sense that if true, it would move your assessment very much, which corresponds to its having *low* (*initial*) *probability*. The same principle applies to deliberative – as opposed to actual – probability. In general, propositions that are highly significant in

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

the sense that they move us much more than their negations have *low* deliberative probability.

Let us apply this idea to some examples. First consider the case of promising to go to the movies. I promised Jim I would go to the movies. Promising is a strong reason to go, but not promising is not a strong reason not to go. So taking account of the fact that I promised moves my decision of whether to go very much toward going. Had I not promised, taking account of the fact that I had not promised would not move my decision very much. Not having promised to go may be viewed as a very weak reason not to go because the fact that I have not promised rules out one of the many reasons to go. Since promising moves me very much and not promising moves me very little, promising has high significance, and not promising has low significance. Correspondingly, promising has low deliberative probability and not promising has high deliberative probability.

The proposition that I promised is a descriptive proposition. The notion of a promise may be value-laden, but that I promised is a contingent event, and it is straightforward to assign probabilities to events such as promisings. We are now thinking about deliberative probabilities, which are hypothetical probabilities assigned in the deliberative position, but we can think of the assignments of deliberative probabilities as being analogous to the assignment of subjective and objective probabilities when we are thinking about contingent propositions.

However, what about more purely normative propositions such as the proposition that it is wrong to kill? Ordinarily, it is difficult to think about what it means to assign probabilities to normative propositions of this sort. Let us nevertheless try to use the same technique as above to determine the deliberative probability of its being wrong to kill.

That it is wrong to kill is a strong reason not to kill, whereas if it were not wrong to kill, then that it is not wrong to kill would not be a strong reason to kill. In general the fact that something is not wrong is a very weak reason to do it; it is only a weak reason to do it because it merely rules out one reason not to do it. Since the wrongness of killing would move me very much in my deliberations but, were it not wrong to kill, the lack of wrongness would not move me very much, that it is wrong to kill is highly significant, and correspondingly it has very low deliberative probability. This shows how in principle deliberative probabilities can be assigned to normative propositions.[29]

[29] This discussion is related to a literature on normative uncertainty, which asks how an agent ought to act when she is uncertain about which moral principles are true (Lockhart 2000; Sepielli 2009; Weatherson 2014; Harman 2015; MacAskill 2016). For further discussion of these relations, see Appendix 2.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

In the application of deliberative probabilities, it is important to make a distinction between *background* and *active* facts. A background fact is a fact that is taken as given in deliberation, and to which we do not assign weight. In contrast, an active fact is a fact explicitly involved in deliberation and to which we do assign weight.

To illustrate this distinction, let R be the proposition that it is wrong to kill. We may treat R as either a background or an active fact. If we treat R as a background fact, then we assign R a deliberative probability of 1; we do not suspend our belief in R, but treat the wrongness of killing as a background condition that is taken as given and not in question. In contrast, if we treat R as an active fact – we want to assess the weight of its being wrong to kill – then because R is highly significant, R receives a deliberative probability very close to zero. In assessing the weights of reasons it is important to distinguish background facts, which are taken as given and not assigned weight, and active facts that are not taken as given and to which we do assign weight.

Finally, why do we need deliberative probability at all? A central idea in this paper is that the weight of a reason for an action a over a default d is the effect that taking the reason into account has on the comparison between a and d. Taking a reason R into account amounts to accepting R as true, and evaluating the comparison between a and d conditional on the truth of R. But what does it mean to not take R into account? Deliberative probabilities are used to model what it is to not take a reason into account.[30] There are several potential ways of modelling the attitude of not taking R into account. The agent might treat R as false or she might adopt an attitude of uncertainty toward R. There are other possibilities: If I do not take the fact that my action will hurt you into account, it might be that the possibility that my action will hurt you has not occurred to me. Let us refer to this attitude as *unawareness* of the possibility. This paper models that attitude that an agent takes towards R when she does not take R into account as uncertainty about R and deliberative probability is used to represent that uncertainty. Observe that the alternative attitude of treating R as false can be approximated by assigning R low deliberative probability. Indeed, as explained above, this is what happens when R is a highly significant proposition. While there are formal models of unawareness (Modica and Rustichini 1994; Dekel *et al.* 1998; Halpern 2001; Heifetz *et al.* 2006), uncertainty is much better understood, and there is a standard and well understood way of modelling it. In addition to modelling the attitude toward R, in my model,

[30] When there are several reasons, deliberative probabilities are used to model what it is to take some but not all of these reasons into account.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

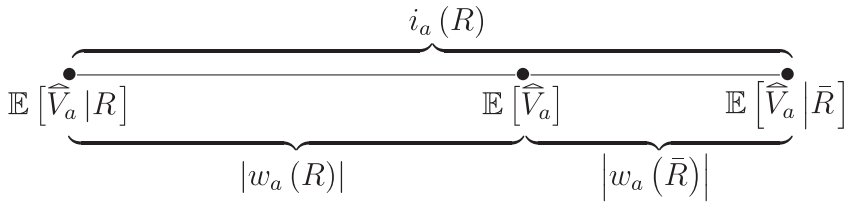


FIGURE 1. Weight and Importance

deliberative probabilities are directly important for modelling what it is to be a reason and the weight of a reason: The expectations in (2) and (3) would not be defined without deliberative probabilities. Deliberative probabilities are also integral to my characterization of independent reasons as can be seen from an examination of Theorem 1 below. These concepts all depend on the attitude that the agent takes toward reasons in the deliberative position.

5.2. A More Formal Discussion

I now discuss the ideas of the preceding section a little more formally. Let us first define the *importance* of reason R for/against a by:

$$i_a(R) = \left|\mathbb{E}\left[\widehat{V}_a \,|R\right] - \mathbb{E}\left[\widehat{V}_a \,|\bar{R}\right]\right|. \tag{8}$$

Importance resembles weight: The importance measures how much difference it makes whether or not R is true for the decision between a and d.

Now consider Figure 1.

The line is a segment of the number line. The expected value $\mathbb{E}[\widehat{V}_a]$ is a real number on the line. $\mathbb{E}[\widehat{V}_a]$ is our initial estimate of the benefit of choosing alternative a rather than the default d. If the agent were to learn R, this would move the expectation in one direction to $\mathbb{E}[\widehat{V}_a \,|R]$. The distance traversed in this movement is the absolute value of the weight of R, $|w_a(R)|$. If the agent were to learn that R does not hold, or in other words, if the agent were to learn the negation $\bar{R}$, this would move her expectation in the opposite direction. Whichever proposition R or $\bar{R}$ was initially viewed as more unlikely would move the agent further from her initial starting point; an unlikely discovery changes one's view more than a likely discovery. Indeed, it follows from the law of iterated expectations that we can relate the probability of R to the weights of R and $\bar{R}$ through

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

the equation[31]

$$P(R) = \frac{1}{1 + \left|\frac{w_a(R)}{w_a(\bar{R})}\right|}. \tag{9}$$

Let us call the ratio $S(R) = \left|\frac{w_a(R)}{w_a(\bar{R})}\right|$ the *deliberative significance* of R. It is not difficult to see that $\left|\frac{w_a(R)}{w_a(\bar{R})}\right| = \frac{1-P(R)}{P(R)}$. So the deliberative significance of R can be interpreted as the odds against R. The greater the odds against R, the more significant it would be to learn that R is true. This accords with the previous informal discussion of significance. Under this interpretation, (9) is a familiar formula for probability in terms of odds.

Observe that $S(R)$ is not indexed by a. More precisely, for all a and b, if $w_a(R) \neq 0$ and $w_b(R) \neq 0$, then $S(R) = \left|\frac{w_a(R)}{w_a(\bar{R})}\right| = \left|\frac{w_b(R)}{w_b(\bar{R})}\right|$. If R is irrelevant to the choice between a and the default – that is, $w_a(R) = 0$ – then equation (9) does not apply.[32]

The significance of R measures how much R moves comparative value relative to its negation. A significant proposition moves comparative value a great deal relative to its negation. Observe that the probability $P(R)$ and the significance $S(R)$ are inversely related: The larger is the significance of R, the smaller is the probability of R.

One can imagine that rather than having intuitions about the deliberative probability of various potential reasons, one has intuitions about the deliberative significance of potential reasons. The examples that I have given above concerning the significance of promising and the significance of its being wrong to kill support this idea. It is then intuitions about deliberative significance that determine deliberative probabilities via (9).

As I said above, the significance of a proposition does not depend on the decision at hand, provided that the proposition is relevant *at all* to the decision at hand (i.e. $w_a(R) \neq 0$). So if R is relevant to the choice between a and the default d, to the choice between b and d, and to the choice between a and b, it is equally *significant* to all. However, it is not equally *important* to all (where importance is defined by (8)). Significance concerns

[31] The law of iterated expectations tells us that

$$\mathbb{E}\left[\widehat{V}_a\right] = \mathbb{E}\left[\widehat{V}_a \,|R\right] P(R) + \mathbb{E}\left[\widehat{V}_a \,|\bar{R}\right](1 - P(R)).$$

Moving $P(R)$ to one side of the equation and all other terms to the other side, doing a little algebra, and appealing to (3) – the definition of $w_a(R)$ – yields (9).

[32] Observe that $w_a(R) = 0$ if and only if $w_a(\bar{R}) = 0$ (provided that both these terms are well defined).

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

a relation between R and its negation as affecting decisions generally, whereas importance relates to the particular decision at hand rather than the relation between R and $\bar{R}$. Indeed for all actions a, $i_a(R) = i_a(\bar{R})$. The importance measures how much difference it makes one way or another whether the R is true to the decision at hand. Weight, significance and importance, are related through the following equation:

$$|w_a(R)| = \frac{i_a(R)}{1 + \frac{1}{S(R)}}. \tag{10}$$

The more significant is R *in general* and the more important is R *to the decision at hand*, the greater (the absolute value of) the weight of R.

If we are to understand deliberative probability, we must also understand *conditional* deliberative probabilities, such as $P(R_2 \mid R_1)$: Once reason R_1 is taken into account, a deliberative probability must be assigned to R_2. How are we to interpret conditional deliberative probabilities? First observe that conditional probabilities are inherent in the underlying deliberative probability measure because $P(R_2 \mid R_1) = \frac{P(R_1 \cap R_2)}{P(R_1)}$. Since deliberative probability is inherent in deliberative significance via equation (9), and conditional probabilities are inherent in probabilities, conditional probabilities are also inherent in deliberative significance – indeed $P(R_2 \mid R_1) = \frac{1+S(R_1)}{1+S(R_1 \cap R_2)}$.[33]

To summarize, one makes judgements about the significance of various propositions and these judgements determine the deliberative probabilities in the deliberative position. These judgements are weaker than judgements of weight. As is made evident by (10), to actually derive judgements of weight, in addition to judgements of general significance, one must also make judgements of the importance of potential reasons for specific decisions.

6. INDEPENDENT VS. NON-INDEPENDENT REASONS

We now return to the problem that was raised in the Introduction: When can the weights of different reasons be added? We saw there that often different reasons stand in complex logical and substantive relations to one another, and a theory of weight must accommodate such relations. If we naively add the weights of different reasons, we may double-count when those reasons have overlapping content. We may also undercount in cases where one reason only gains its force in the presence of another reason. We have provided a framework for keeping track of these interrelations by

[33] One can update deliberative significance directly via the equation $S(R_2 \mid R_1) = \frac{1+S(R_1 \cap R_2)}{1+S(R_1)} - 1$. $S(R_2 \mid R_1)$ is the deliberative significance of R_2 conditional on R_1. Intuitively, $S(R_2 \mid R_1)$ measures the significance of R_2 once R_1 has already been taken into account.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

positing that reasons have the structure of conditional expectations; that is, the weight of a reason is $\mathbb{E}[V_a \,|R] - \mathbb{E}[V_a]$. $\mathbb{E}[V_a \,|R]$ is a conditional expectation that takes R as an argument, and $\mathbb{E}[V_a]$ is a constant that does not depend on R. The structure of conditional expectations – specifically the relations imposed on interrelated propositions – is well understood and we shall explore this structure in more depth in Section 7.

To address the question of when weights can be added, this section does several things. First, it examines the limiting case in which reasons are independent so that weights can be added without either under- or over-counting. Then it defines *conditional weights*, and shows that even in the absence of independence, it is always valid to add certain conditional weights in order to decide what to do. Finally, it returns to the examples of interrelated reasons with which the paper opened, and shows how they can be handled in the paper's framework.

6.1. Independent Reasons

Let $\mathscr{R} = \{R_1, \ldots, R_n\}$ be a collection of reasons. Below, I will assume that your reasons are mutually consistent. More precisely, say that $\mathscr{R}$ is *consistent* if $P\left(\bigcap_{i=1}^{n} R_i\right) > 0$. Throughout this section, whenever discussing a collection of reasons $\mathscr{R}$, I will assume that $\mathscr{R}$ is consistent. A reason R_0 is *consistent with* $\mathscr{R}$ if $\mathscr{R} \cup \{R_0\}$ is consistent.

We can extend the notion of the weight of a reason to collections of reasons. In particular, define the weight of a consistent collection of reasons $\mathscr{R} = \{R_1, \ldots, R_n\}$ for/against a as:

$$\hat{w}_a(\mathscr{R}) = w_a\left(\bigcap_{i=1}^{n} R_i\right). \tag{11}$$

In other words, the weight of a collection of reasons is just the weight of their conjunction (or, more formally, their intersection). The hat $\hat{}$ in $\hat{w}_a(\mathscr{R})$ differentiates weight applied to a collection of reasons from weight applied to a single reason $w_a(R)$. Using the convention $\bigcap \emptyset = \Omega$, observe that $\hat{w}_a(\emptyset) = w_a(\Omega) = 0$.[34] In other words, the weight of a collection of reasons containing nothing is zero; there are no reasons in the collection to move our assessment of the decision.

We can also define *conditional weights*. That is, we can define the weight of a reason conditional on having accepted some other reasons. This is the weight of the reason not from the initial deliberative position in which knowledge of all relevant reasons has been suspended, but from an intermediate deliberative position in which some of these reasons have

[34] This follows from the fact that $\mathbb{E}[\widehat{V}_a \,|\Omega] = \mathbb{E}[\widehat{V}_a]$, so that $w_a(\Omega) = \mathbb{E}[\widehat{V}_a \,|\Omega] - \mathbb{E}[\widehat{V}_a] = 0$. See also the *No Information* axiom in Section 7.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

been accepted again. The *weight of reason R_0 for/against a conditional on $\mathscr{R}$* (where $\mathscr{R} = \{R_1, \ldots, R_n\}$) is:

$$w_a\left(R_0 \,|\mathscr{R}\right) = \mathbb{E}\left[\widehat{V}_a \,\middle|\, R_0 \cap \left(\bigcap_{i=1}^{n} R_i\right)\right] - \mathbb{E}\left[\widehat{V}_a \,\middle|\, \bigcap_{i=1}^{n} R_i\right]. \tag{12}$$

That is to say, $w_a\left(R_0 \,|\mathscr{R}\right)$ is the difference in the expected comparative value of a that conditioning on R_0 makes when one has already conditioned on the reasons in $\mathscr{R}$. It follows from our definitions that

$$w_a\left(R_0 \,|\mathscr{R}\right) = \hat{w}_a\left(\{R_0\} \cup \mathscr{R}\right) - \hat{w}_a\left(\mathscr{R}\right). \tag{13}$$

In Section 7, I discuss the possibility of taking weight as fundamental, rather than as derived from values and deliberative probabilities. If this is done, then (13) can be taken as the *definition* of conditional weights.[35] This shows that in the same way that conditional probabilities can be derived from unconditional probabilities, conditional weights can be derived from unconditional weights.

As a matter of notation, for any potential reasons R_0 and R_1, I write $w_a(R_0 \,|\, R_1)$ instead of $w_a(R_0 \,|\, \{R_1\})$, so that, in particular, $w_a\left(R_0 \,|\mathscr{R}\right) = w_a\left(R_0 \,\middle|\, \bigcap_{i=1}^{n} R_i\right)$. This notational convention only has potential for ambiguity if the reason R_1 that one conditions on is $R_1 = \varnothing$; however since I exclude inconsistent reasons $w_a(R \,|\, \varnothing)$ means that weight of R is conditional on an empty set of reasons, not conditional on the inconsistent reason. So $w_a(R \,|\, \varnothing) = w_a(R \,|\, \Omega) = w_a(R)$: If the set of reasons that one conditions on is empty, one remains with one's initial assessment. Finally $w_{a,b}\left(R_0 \,|\mathscr{R}\right)$ is defined analogously to $w_a\left(R_0 \,|\mathscr{R}\right)$, with $w_{a,\,b}$ playing the role of w_a.

The importance of the notion of a conditional weight stems from the fact that in general the weight of a reason changes when we take other reasons into account. That is to say, in general $w_a(R_1) \neq w_a(R_1 \,|\, R_2)$. Indeed, it is possible that $w_a(R_1) > 0$, so that R_1 is a reason for action a, but $w_a(R_1 \,|\, R_2) < 0$, which means that conditional on R_2, R_1 is a reason against a. So reasons can change their valence depending on the context, which is here modelled by the set of other reasons that have already been taken into account. Thus, when asking whether R_1 is a reason for a, it is important to specify the context. Independence of reasons occurs precisely when the weight of a reason is unaffected by other reasons. This is captured formally by the following definition.

Definition 3. *$\mathscr{R}$ is a collection of* independent reasons *for/against* a *if and only if for all reasons* R_0 *in $\mathscr{R}$ and for all subsets $\mathscr{S}$ of $\mathscr{R}$ with* R_0 *not in $\mathscr{S}$,*

$$w_a\left(R_0 \,|\mathscr{S}\right) = w_a(R_0).$$

To abbreviate, we also say that $\mathscr{R}$ is a-independent.

[35] This requires of course that the function $\hat{w}_a$ is defined in terms of the function w_a via (11).

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

In other words, $\mathscr{R}$ is a collection of independent reasons if conditioning on any subset of reasons $\mathscr{S}$ does not alter the weight of any reason in $\mathscr{R}$ outside the set $\mathscr{S}$. The following proposition is immediate from definitions.

Proposition 3. *The collection of reasons $\mathscr{R}$ is* a-*independent if and only if for every subset $\mathscr{S}$ of $\mathscr{R}$, $\hat{w}_a(\mathscr{S}) = \sum_{R \in \mathscr{S}} w_a(R)$.*

In other words, for a collection of independent reasons the weight of any subcollection is equal to the sum of the weights of the individual reasons. Thus when weights are independent, they can be added.

Define a collection of reasons $\mathscr{R} = \{R_1, \ldots, R_n\}$ as (a, b)-*independent* if $w_{a,b}(R_0 | \mathscr{S}) = w_{a,b}(R_0)$ for all $R_0 \in \mathscr{R}$ and all non-empty $\mathscr{S} \subseteq \mathscr{R}$ with $R_0 \notin \mathscr{S}$. This is identical to the definition of a-independence, except that $w_{a,b}$ has been substituted for w_a. Observe that the conjunction of a-independence and b-independence is a sufficient but not necessary condition for (a, b)-independence. Let us now interpret $\mathscr{R} = \{R_1, \ldots, R_n\}$ as all of the agent's actual reasons. Recall the notion of preference on the basis of one's reasons $\succsim$ defined by (7) in Section 4. Recall also that *initial neutrality* is the assumption that for all actions a and b, $\mathbb{E}[V_a] = \mathbb{E}[V_b]$. Combining the analogue of Proposition 3 for (a, b)-independence with Proposition 1 and (7) yields the following result.

Proposition 4. *Assume initial neutrality. Assume moreover that $\mathscr{R}$ is* (a, b)-*independent. Then the following three conditions are equivalent: (1)* a$\succsim$b*; (2)* $\mathbb{E}\left[V_a \left| \bigcap_{i=1}^n R_i \right.\right] \geq \mathbb{E}\left[V_b \left| \bigcap_{i=1}^n R_i \right.\right]$*; (3)* $\sum_{i=1}^n w_{a,b}(R_i) \geq 0$.

In words, a is preferred to b if and only if the expected value of a is greater than the expected value of b conditional on one's reasons if and only if the sum of weights of reasons in favour of a over b is greater than the sum of weights of reasons in favour of b over a; to understand the last condition, observe that if $R_1, \ldots R_k$ support a over b (that is, if R_i is among these reasons, $w_{a,b}(R_i) > 0$) and $R_{k+1}, \ldots, R_n$ support b over a (so that if R_j is among these reasons, $w_{a,b}(R_j) < 0$, so that $w_{b,a}(R_j) = -w_{a,b}(R_j) > 0$), then (3) is equivalent to $\sum_{i=1}^k w_{a,b}(R_i) \geq \sum_{j=k+1}^n w_{b,a}(R_j)$.

In Section 6.4, I will characterize independent reasons in terms of the structure of underlying values and deliberative probabilities. This will give greater insight into when reasons are independent. However, first I will compare independent to non-independent reasons with regard to the aggregation of reasons in practical decision-making.

6.2. Non-independent Reasons and Conditional Weights

If reasons are not independent, we cannot add the weights of reasons. However, we can still add *conditional* weights. Suppose for example that there are three reasons R_1, R_2 and R_3. If the reasons are a-independent,

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

then, as we have seen in the previous section, the weight of the conjunction is the sum of the *unconditional* weights of the individual reasons:

$$w_a(R_1 \cap R_2 \cap R_3) = w_a(R_1) + w_a(R_2) + w_a(R_3). \tag{14}$$

Equation (14) is *not* valid for non-independent reasons. However something similar is valid. For non-independent reasons, the weight of the conjunction is the sum of the *conditional* weights of the individual reasons:

$$w_a(R_1 \cap R_2 \cap R_3) = w_a(R_1) + w_a(R_2 \,|\, R_1) + w_a(R_3 \,|\, R_1 \cap R_2).$$

In general, to form the weight $w_a\left(\bigcap_{i=1}^{n} R_i\right)$ of the conjunction of all of our reasons, we can perform the following procedure.

1. Start with the unconditional weight $w_a(R_1)$,
2. then add the weight $w_a(R_2 \,|\, R_1)$ of R_2 conditional on R_1,
3. then add the weight of $w_a(R_3 \,|\, R_1 \cap R_2)$ of R_3 conditional on R_1 and R_2,

   $\cdots$,

n. and finally add the weight $w_a\left(R_n \left| \bigcap_{i=1}^{n-1} R_i \right.\right)$ of R_n conditional on all preceding reasons to arrive at the aggregate weight of all reasons for or against a.

We terminate with the sum:

$$w_a(R_1) + w_a(R_2 \,|\, R_1) + w_a(R_3 \,|\, R_2 \cap R_2) + \cdots + w_a\left(R_n \left| \bigcap_{i=1}^{n-1} R_i \right.\right).$$

When adding non-independent reasons, we must condition on increasingly more information as we add more reasons. The reasons accumulate as we go. We must assess the weight of each new reason in light of the weights of the reasons that we have already taken into account. The discussion is summarized in Proposition 5, which is the analogue of Proposition 3 for independent reasons.

Proposition 5. *For any collection of reasons* $\mathrm{R}_1, \ldots, \mathrm{R}_n$,[36]

$$w_a\left(\bigcap_{i=1}^{n} R_i\right) = \sum_{i=1}^{n} w_a\left(R_i \left| \bigcap_{j=1}^{i-1} R_j \right.\right).$$

The analogue of Proposition 4 for non-independent reasons is the following.

[36] Recall that $\bigcap_{i=1}^{0} R_i = \Omega$ and $w_a(R_1 \,|\, \Omega) = w_a(R_1)$.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Proposition 6. *Assume initial neutrality. Then the following three conditions are equivalent: (1)* a$\gtrsim$b; *(2)* $\mathbb{E}\left[V_a \left|\bigcap_{i=1}^{n} R_i\right.\right] \geq \mathbb{E}[V_b \left|\bigcap_{i=1}^{n} R_i\right.]$;
(3) $\sum_{i=1}^{n} w_{a,b}\left(R_i \left|\bigcap_{j=1}^{i-1} R_j\right.\right) \geq 0$.

That is, to select an action when reasons are not independent, we add conditional weights.

A final note is in order. Section 4 discussed the possibility that not all reasons are relevant to all decisions. Let $\mathscr{R}$ be the set of all your reasons that are relevant to any decision between any pair of actions. Say that a subset $\mathscr{S}$ of $\mathscr{R}$ is (a, b)-*complete* if for all $\{R_{i_1}, \ldots, R_{i_h}\} \subseteq \mathscr{R}$, $w_{a,b}(\bigcap_{j=1}^{h} R_{i_j} | \mathscr{S}) = 0$. This means that once $\mathscr{S}$ is taken into account, no further reasons have any weight. Say that $\mathscr{S}$ is *minimal* (a, b)-complete if no proper subset of $\mathscr{S}$ is (a, b)-complete. Once a minimal (a, b)-complete set reasons has been taken into account, no other reasons remain relevant to the decision between a and b. Suppose for example that $\{R_1, \ldots, R_k\}$ with $k < n$ is a minimal (a, b)-complete set. Then $\sum_{i=1}^{n} w_{a,b}(R_i \left|\bigcap_{j=1}^{i-1} R_j\right.) = \sum_{i=1}^{k} w_{a,b}(R_i \left|\bigcap_{j=1}^{i-1} R_j\right.)$, so that (under initial neutrality) we should take action a over b if and only if $\sum_{i=1}^{k} w_{a,b}(R_i \left|\bigcap_{j=1}^{i-1} R_j\right.) \geq 0$; in other words, we need not consider reasons relevant only to other decisions.

6.3. Examples from the Introduction

I now revisit some of the examples from the Introduction to show how they would be dealt with in the framework that I have developed. Recall the example in which I am trying to decide whether to take a treatment for a disease. Two relevant reasons to take the treatment may be:

- R_1 : The treatment would prolong my life by at least 1 year.
- R_2 : The treatment would prolong my life by at least 2 years.

Observe that $R_2 = R_1 \cap R_2$. Clearly these reasons overlap. Handling overlapping reasons such as this was one of the main motivations for developing the framework of this paper.

Let a be the action corresponding to taking the treatment, and let the default d be the action of not taking the treatment. Clearly, the weight of R_2 in favour of a is greater than the weight of R_1 in favour of a. To be concrete, suppose that $w_a(R_1) = 3$ and $w_a(R_2) = 5$. Then observe that

$$w_a(R_2 | R_1) = w_a(R_1 \cap R_2) - w_a(R_1) = 5 - 3 = 2,$$
$$w_a(R_1 | R_2) = w_a(R_1 \cap R_2) - w_a(R_2) = 5 - 5 = 0.$$

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

As a special case of Proposition 5, it now follows that

$$w_a(R_1 \cap R_2) = w_a(R_1) + w_a(R_2|R_1) = 2 + 3 = 5$$
$$= w_a(R_2) + w_a(R_1|R_2) = 5 + 0 = 5.$$

To arrive at the total weight, I add appropriate conditional weights. The order in which the reasons are considered does not matter. If first I consider that the treatment will prolong my life by at least one year, I add the weight $w_a(R_1) = 3$, but having taken this into account, considering that the treatment will actually prolong my life by at least two years, I add a weight of $w_a(R_2 \mid R_1) = 2$, arriving at a total weight of 5. This second reason was partially – but not fully – taken into account when I considered the first reason. On the other hand, if first I consider that the treatment will prolong my life by at least two years, then this has a weight of $w_a(R_2) = 5$, and then considering that the treatment will prolong my life by at least one year does not amount to adding a new significant reason – that reason has already been taken into account: We have $w_a(R_1 \mid R_2) = 0$. In either order of considering the reasons, the total weight of their conjunction is 5. In sum, the framework naturally handles the overlap between reasons, and is not subject to any problem of double-counting when reasons are considered in conjunction.

Let us now consider a more conceptually challenging example, which was also raised in the introduction. Suppose that action a is the action of going to the movies, and the default d is the action of staying at home. Let us assume for simplicity that $V_d \equiv 0$. That is the default of staying at home always (i.e. in every state) yields a value of zero.

Let us suppose that before taking reasons into account, it would be better to stay at home. This violates the initial neutrality assumption but (as explained in footnotes 25–26) the initial neutrality assumption is inessential and nothing of import changes if it is eliminated. Our assumption about the value of the default – that in the absence of reasons, it would be better to stay at home – amounts to the assumption that $\mathbb{E}[V_a] < 0$. Now consider the following reasons (from the Introduction):

R_1 : I promised Jim I would go.
R_2 : It is important to keep one's promises.

The reason R_2 has high deliberative significance conditional on R_1; similarly, R_1 has high deliberative significance conditional on R_2. Using the inverse relation between deliberative significance and deliberative probability, this is equivalent to assuming that $P(R_2 \mid R_1)$ and $P(R_1 \mid R_2)$ are small. Assume in fact that the deliberative significance is high enough that these two probabilities are approximately zero. Suppose that V_a is positive when and only when R_1 and R_2 are both true simultaneously;

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

that is, it would be better to go to the movies than to stay home if it is true both that I promised that I would go to the movies and that it is important to keep one's promises. Suppose moreover that V_a has a roughly constant negative value when $R_1 \cap R_2$ is false; that is, given that one of the propositions R_1 and R_2 is false, it does not matter very much whether the other is. For example, it may be that Jim would be equally happy if I were to go to the movies (and equally unhappy if I were not to go) regardless of whether I promised to go, so that promising by itself does alter the comparison between going and not going from the standpoint of Jim's happiness.[37] Then if $P(R_1 \mid R_2)$ and $P(R_2 \mid R_1)$ are small enough, it follows that $\mathbb{E}[V_a \mid R_1] \approx \mathbb{E}[V_a \mid R_2] \approx \mathbb{E}[V_a] < 0$. On the other hand our assumptions also imply that $\mathbb{E}[V_a \mid R_1 \cap R_2] > 0$. It then follows that $w_a(R_1) \approx w_a(R_2) \approx 0$, but $w_a(R_1 \mid R_2) \approx w_a(R_2 \mid R_1) \approx w_a(R_1 \cap R_2) > 0$. That is, if it were not important to keep one's promises, then the fact that I made a promise to Jim would not be a reason to go, and if I had not made a promise to Jim, then the fact that it is important to keep one's promises would not be a reason to go. Conditional on either reason, the other reason carries the full weight of their conjunction. The two reasons are, in a sense, 'perfect complements'. In this case, it is not possible to divide the weight of the conjunction among the two reasons in any meaningful way; it is only in conjunction that the two reasons have weight. This seems true in this particular example. There are many examples in which the weight can only be assigned jointly to a set of reasons; if any subset is put into the *background*, the remaining *active* reasons carry all of the weight (see Section 5.1 for the distinction between background and active reasons). Thus in different contexts, in which different background conditions are taken to be assumed, we may assign the weights to different reasons. There are other less extreme cases in which some subsets of reasons have weight on their own, but the whole is still more than the sum of its parts. That is, $w_a(R_1) > 0$, $w_a(R_2) > 0$ and $w_a(R_1 \cap R_2) > w_a(R_1) + w_a(R_2)$. The model can easily represent such possibilities.

6.4. Characterization of Independent Reasons in terms of Values and Probabilities

We can characterize independent reasons in terms of properties of the corresponding value functions and deliberative probabilities. To do so, we actually need a stronger definition of independence. For any reason R, let $R^1 = R$; that is, R^1 is just the reason R itself, and let $R^0 = \bar{R}$; so R_0 is the negation (formally, the complement) of R. We can refer to R^1 as the

[37] Since we are assuming that $\mathbb{E}[V_a] < \mathbb{E}[V_d] = 0$, the effect on Jim's happiness by itself is not enough to justify going to the movies all things considered if R_1 and R_2 are not taken into account.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

'affirmation' of R and R^0 as the 'disaffirmation' of R. Let $\mathscr{R} = \{R_1, \ldots, R_n\}$, and define a *truth assignment* for $\mathscr{R}$ to be a function $t : \{1, \ldots, n\} \to \{0, 1\}$. Then define $\mathscr{R}^t = \left\{R_1^{t(1)}, \ldots, R_i^{t(i)}, \ldots, R_n^{t(n)}\right\}$ to be such that for each potential reason R_i, $R^{t(i)}$ affirms R_i if $t(i) = 1$ and disaffirms R_i if $t(i) = 0$. For example if $\mathscr{R} = \{R_1, \ldots, R_5\}$ and $t(i) = 1$ if i is odd and $t(i) = 0$ if i is even, then $\mathscr{R}^t = \left\{R_1, \bar{R}_2, R_3, \bar{R}_4, R_5\right\}$. Say that $\mathscr{R}$ is *logically independent* if for all truth assignments t, $\mathscr{R}^t$ is consistent.[38] We now have the following strengthened definition of independence.

Definition 4. *Let $\mathscr{R}$ be a collection of logically independent reasons. $\mathscr{R}$ is a collection of* strongly independent reasons *for/against* a *if and only if for all truth assignments* t, *$\mathscr{R}^t$ is* a*-independent. To abbreviate, we say that $\mathscr{R}$ is* strongly *a*-independent.

Thus, a collection of reasons is strongly a-independent if the collection is a-independent and negating any subset to the reasons, the collection remains a-independent.[39] I need one more definition: *$\mathscr{R}$ contains an a-empty reason* if there exists R in $\mathscr{R}$ such that for all subsets $\mathscr{S}$ of $\mathscr{R}$ with R not in $\mathscr{S}$, $w_a(R \mid \mathscr{S}) = 0$. An a-empty reason is a reason that has no weight, no matter which other reasons we condition on. I now state a theorem that characterizes strong a-independence in terms of properties of the comparative value of a, $\widehat{V}_a$, and the deliberative probability P.

Theorem 1. *Suppose that $\mathscr{R}$ is logically independent and contains no* a*-empty reasons. Then the following two statements are equivalent:*

1. *$\mathscr{R}$ is strongly* a*-independent.*
2.
   (a) $\mathrm{R}_1, \ldots, \mathrm{R}_\mathrm{n}$ *are mutually probabilistically independent with respect to* P.
   (b) *There exist random variables* U_i, $\mathrm{i} = 1, \ldots, \mathrm{n}$ *such that*
      i. $\widehat{V}_a = \sum_i U_i$, *and*
      ii. $\mathbb{E}\left[U_i \middle| \bigcap_{j=1}^n R_j^{t(j)}\right] = \mathbb{E}\left[U_i \middle| R_i^{t(i)}\right]$ *for all* i *and all truth assignments* t.

[38] For the formal definition of consistency, see the beginning of Section 6.1. The definition here differs from the standard definition of logical independence. However, logical independence has the right spirit: The idea is that any combination of the propositions could be true while the others are false.

[39] The reader may observe that for probability measures – as opposed to weight functions – the conditions of independence and strong independence are equivalent. In other words, let $R_1, \ldots, R_n$ be independent events according to the probability measure μ, then if we take the complement of any subset of these events arriving at a collection such as $\bar{R}_1, R_2, \bar{R}_2, \ldots, \bar{R}_{n-1}, R_n$, the resulting collection must also be independent according to μ. However, the weight function w_a is not a probability measure, and strong a-independence is indeed a property that is harder to satisfy than a-independence.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Let us interpret the result. What is needed in order for reasons to be strongly independent? First, the reasons $R_1, \ldots, R_n$ must be probabilistically independent with respect to deliberative probability. That is, in deliberation, learning that any one of the reasons obtains provides no information about any of the other reasons. Second, the (comparative) value of a must be decomposable into a sum of factors U_i (condition 2(b)i), and each factor U_i is influenced (in expectation) by only one reason R_i (condition 2(b)ii).

Let me make this concrete by means of an example. This is an important example of independent reasons. Suppose that I am a utilitarian and the various actions a correspond to different policies. $U_i^a(\omega)$ is the utility of agent i in state ω if action a is taken. There is a default action d such that $U_i^d \equiv 0$ for all i; in other words, the default action produces a zero utility at any state of the world. Perhaps the action a is some policy contemplated by the government and the default d is doing nothing or sticking with the status quo. For a utilitarian, $\widehat{V}_a = V_a = \sum_i U_i^a$. That is, since the default delivers zero utility, the benefit (or cost) that results from choosing a is the sum of utilities this choice would lead to for the people affected. Let R_i^a be the proposition that action a is good (or bad) for agent i. (This proposition can be made more precise in various ways.) Suppose the fact that the policy is good for one agent is completely uninformative about its effects on any other agent. This is understood from the standpoint of the deliberative position: When I deliberate about whether the action is taken, I do not take the proposition that it is good for any given agent to be informative with regard to the question of whether it is good for any other agent. Then according to Theorem 1, the reasons R_i^a are independent reasons for or against taking the action. Thus utilitarianism is a paradigmatic case of independent reasons, where the reasons are the net utility effects that the action under consideration has on various people.

Finally, I mention one variant of Theorem 1. Let $\mathscr{F}_{\mathscr{R}}$ be the field generated by $\mathscr{R}$.[40] For any event R, let 1_R be the indicator function that takes value 1 at states ω in R and takes value 0 in states ω not in R. Then in Theorem 1, condition 2b can be replaced by[41]

$$\mathbb{E}\left[\widehat{V}_a \,|\mathscr{F}_{\mathscr{R}}\right] = \sum_{i=1}^{n} w_a(R_i)\,1_{R_i} + \sum_{i=1}^{n} w_a\left(\bar{R}_i\right)1_{\bar{R}_i} + \mathbb{E}\left[\widehat{V}_a\right]. \qquad (15)$$

[40] This means that $\mathscr{F}_{\mathscr{R}}$ is the closure of $\mathscr{R}$ under intersection and complementation; in this paper, I use the words field and algebra interchangeably.

[41] For a formal treatment of what it means to condition a random variable on a σ-field, see Section 34 of Billingsley (1995). For comparison with the definition there, note that the field $\mathscr{F}_{\mathscr{R}}$ is also a σ-field because $\mathscr{R}$ is finite.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Note that this is still a characterization in terms of V and P, because $w_a(R_i)$ and $w_a\left(\bar{R}_i\right)$ are defined in terms of V and P (see Definition 2). This has the advantage of eliminating the existential quantifier in condition 2(b)ii by effectively identifying U_i: that is, $U_i = w_a\left(R_i\right)1_{R_i} + w_a\left(\bar{R}_i\right)1_{\bar{R}_i}$ for $i = 1, \ldots, n-1$ and $U_n = \widehat{V}_a - \sum_{i=1}^{n-1} U_i$.[42] The condition (15) says that we start with $\mathbb{E}\left[\widehat{V}_a\right]$ and for each reason R_i and each state ω, we add $w_a(R_i)$ or $w_a\left(\bar{R}_i\right)$, depending on whether R_i or $\bar{R}_i$ obtains. Observe that the probabilistic independence of $R_1, \ldots, R_n$ (condition 2a) is still required.

7. TAKING WEIGHT AS FUNDAMENTAL

This paper has analysed weight in terms of value and deliberative probability. But what is really more fundamental: value or weight? This is a substantive philosophical question. The view that the weight of reasons derives from values is a natural one and is harmonious with consequentialism. For such views, the value of outcomes can be broadly conceived.[43] Scanlon (1998) defends an opposing buck-passing account of value that takes reasons to be fundamental: '[B]eing good, or valuable, is not a property that itself provides a reason to respond to a thing in certain ways. Rather, to be good or valuable is to have other properties that constitute such reasons.' In this section, I will show that the model developed in this paper can accommodate either view, and in particular can accommodate the view that weight is more fundamental than value. In the process of elaborating on this, I will describe the structure of weight more explicitly than I have done above. To summarize, this section addresses two questions of fundamental importance for a theory of weight:

1. What is more fundamental: weight or value?
2. What is the structure of weight?

To see how the model presented here is compatible with taking weight as fundamental, suppose that we start with the weight $w_a(R)$ of each potential reason R for action a (over the default d). In this section, I show that if we impose certain axioms on the weight function $w_a(R)$, then we can derive implicit values V_a and deliberative probabilities P from weight. These implicit values and probabilities are related to weight as in (3). However, rather than treating (3) as a *definition* of weight in terms of value and deliberative probability, we can treat (3) as establishing a

[42] This is one way of choosing the U_i so that condition 2b of the theorem holds; there are also other ways.

[43] Maguire (2016) presents a value-based theory of reasons that analyses what it is to be a reason in terms of value promotion.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

structural relationship between weight and the values and deliberative probabilities that are implicitly defined by weights. The theory of weight can be grounded in the axioms of weight. When these axioms are satisfied, it is *as if* the weights are derived by updating on reasons in the deliberative position. Or alternatively we can take the deliberative position, values and deliberative probabilities to be more fundamental and then this section shows that weights derived in the deliberative position satisfy – indeed are characterized by – certain axioms.

To summarize, under the axioms, we can take weight as fundamental and treat values as derived. Or, as in the preceding sections of the paper, we can take values and deliberative probabilities as fundamental, and derive weights. So either weights or values can be taken as fundamental and the other derived. Ultimately, whether weights or values are fundamental is a substantive philosophical question. What I show here is that the model of this paper does not prejudge the question. It is consistent with either point of view.

The axiomatization to be presented in this section provides an answer to the second question above, viz., what is the structure of weight? In Section 3.2, I encapsulated the goal of this paper as being 'to develop a theory of weight on a par with probability theory that details the structure of weight, and specifically the relations among weights of interrelated propositions.' In the theory that I have presented, weight does not have precisely the structure of probability; rather weight has the structure of *conditional expectations*. Conditional expectations are related to probability measures but they satisfy different axioms than do probability measures.[44]

The axioms of this section impose the structure of conditional expectations on weight. Indeed, the structure of conditional expectations is well understood. Bolker (1966) axiomatized conditional expectations. Bolker (1967) and Jeffrey (1990) applied this structure to study the *desirability* of propositions.[45] Weight differs in interpretation: it is a *double difference* in values.[46] My axiomatization of weight differs from Bolker's axiomatization of conditional expectations in several ways. I will mention two of the most important ways. First, I work with a finite algebra of propositions (modelled as sets of states) rather than with a complete atomless Boolean algebra.[47] This difference is motivated

[44] Indeed, a conditional expectation is a quotient, whose numerator is a signed measure and whose denominator is a probability measure.

[45] Jeffrey (1990) presents formulas similar to (9). Jeffrey's formula expresses probability in terms of desirability instead of, as in (9), in terms of weight.

[46] The values of which take the double difference might or might not be interpreted as 'desirabilities'.

[47] Domotor (1978) presented an axiomatization of Bolker-Jeffery decision theory for finite algebras. Unlike Domotor, I do not axiomatize a preference on propositions that

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

by considerations of simplicity and transparency.[48] Second, my *Common Ratios* axiom doesn't concern the structure of conditional expectations, but concerns rather the relation among the weights of a given reason (and of its complement) for different actions.

Let $\mathscr{P}$ be the set of potential reasons. In this section, I will assume that $\mathscr{P}$ is the set of all finite subsets of a finite set Ω. This assumption is artificial, but it simplifies the analysis without modifying the main ideas in any essential way. To reduce verbiage, in this section I will refer to potential reasons simply as reasons. Ω and $\varnothing$ are referred to as *trivial* reasons.[49] All other reasons are *non-trivial*. Two reasons R_1 and R_2 are *disjoint* if $R_1 \cap R_2 = \varnothing$. In other words, disjoint reasons are reasons that are mutually inconsistent.

Let A be the set of actions other than the default.[50] A *weight function* is a function $w : A \times \mathscr{P} \to \mathbb{R}$ that maps an action a and a potential reason R to a real number. To match my previous notation, I write $w_a(R)$ instead of $w(a, R)$. To start, I assume the following conditions, each of which apply for all actions a (other than the default).[51]

Simplification. For all pairs of disjoint reasons R_1 and R_2,

$$w_a(R_1) \neq w_a(R_2).$$

No Information. $w_a(\Omega) = 0$.

Averaging. For all pairs of disjoint non-trivial reasons R_1 and R_2,

$$w_a(R_1) < w_a(R_2) \Rightarrow w_a(R_1) < w_a(R_1 \cup R_2) < w_a(R_2).$$

Simplification says that no two disjoint reasons have exactly the same weight. This does rule out important phenomena: In particular, together with the other assumptions, it rules out the possibility that some reasons are irrelevant to some decisions; that is, it rules out that for some action a and non-trivial reason R, $w_a(R) = 0$. However notice that any weight function not satisfying *Simplification* can be arbitrarily well approximated by one that does. In particular, we can have $w_a(R) \approx 0$, where $w_a(R)$ is arbitrarily close to zero. The purpose of this assumption is to simplify the analysis.

represents expectations conditional on those propositions; I axiomatize weights – which are conditional expectations – directly.

[48] In the context of Bolker-Jeffrey decision theory, the atomlessness of the underlying algebra can be motivated by the existence of a randomization device such as a roulette wheel, or a coin that can be flipped indefinitely many times. In the current context, in which probability is interpreted as deliberative probability, such devices are less natural.

[49] Perhaps the word 'trivial' is not apt for $\varnothing$ since it is impossible for $\varnothing$ to occur.

[50] The default is not included because the weight of R for or against the default is always zero, since regardless of R, the default is neither better nor worse than itself.

[51] For the default, see footnote 50.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

The *No Information* assumption says that the trivial reason Ω has no weight. This is natural: Since Ω contains no information,[52] to take Ω into account has no effect on decisions and so carries no weight.

The averaging assumption is due to Bolker (1966).[53] It is one of the axioms that Bolker used to characterize conditional expectations. It is an intuitive axiom for reasons. It says that the weight of a disjunction of disjoint reasons is in between their individual weights. For example, when considering going to the movies, the weight of the proposition that either I promised I would go or I didn't promise but Jim would enjoy it if I went is intermediate between the weight of the proposition that I promised I would go and the weight of the proposition that I didn't promise but Jim would enjoy it if I went. It would be odd if the weight of the disjunction were either greater or less than the weights of both the disjuncts.

When reasons R_1 and R_2 are disjoint, *Averaging* tells us that the weight $w_a(R_1 \cup R_2)$ is between $w_a(R_1)$ and $w_a(R_2)$, but it tells us nothing about precisely where in the interval between $w_a(R_1)$ and $w_a(R_2)$, $w_a(R_1 \cup R_2)$ lies, nor is there any relation posited between where in the interval between $w_a(R_1)$ and $w_a(R_2)$, $w_a(R_1 \cup R_2)$ lies, and where in the interval between $w_a(S_1)$ and $w_a(S_2)$, $w_a(S_1 \cup S_2)$ lies, where R_1, R_2, S_1 and S_2 may be logically related reasons. Bolker (1966) imposed further constraints via assumption he termed *Impartiality*, but for the combination of assumptions to achieve their desired end, Bolker required a strong richness assumption on the class of propositions.[54] To keep technicalities to a minimum and also because the richness assumption may be unnatural in the current setting,[55] I do not adopt the richness assumption here.

Consider two reasons R and S such that $S \neq \emptyset$ and $R \subseteq S$. $S - R$ is the relative complement of R in S. It follows from *Simplification* and *Averaging* that $w_a(S)$ is between $w_a(R)$ and $w_a(S - R)$. Moreover, it follows from our assumptions that there is a unique number $\rho_a^{R|S}$ between 0 and 1, such that

$$\rho_a^{R|S} w_a(R) + \left(1 - \rho_a^{R|S}\right) w_a(S - R) = w_a(S). \tag{16}$$

Indeed, $\rho_a^{R|S} = \frac{w_a(S) - w_a(S-R)}{w_a(R) - w_a(S-R)}$.[56] To further discipline averaging, I impose a condition that relates $\rho_a^{R|S}$ for different pairs of reasons R, S:[57]

[52] Ω contains no information beyond background information that has already been taken into account.

[53] Bolker's condition also contains a clause in which $<$ is replaced by $=$. This clause holds vacuously in our setting because of our other assumptions.

[54] Bolker (1966) worked in a complete atomless Boolean algebra.

[55] See the reference to a randomization device in footnote 48.

[56] *Simplification* and the fact that $S \neq \emptyset$ imply that the denominator is not equal to zero. Observe that in the special case where $R = S$, $\rho^{R|S} = 1$.

[57] Both *Chain Rule* and *Order Independence* below, like the other axioms above, are assumed to hold for all actions a other than the default.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Chain Rule. For all reasons R, S and T such that $S \neq \emptyset$ and $R \subseteq S \subseteq T$,

$$\rho_a^{R|T} = \rho_a^{R|S} \rho_a^{S|T}.$$

To get a handle on the meaning of *Chain Rule*, consider the following alternative condition.

Order Independence. Let R_1, R_2 and R_3 be pairwise disjoint reasons, and let α_1 and α_2 be numbers such that

$$\alpha_1 w_a(R_1) + (1-\alpha_1) w_a(R_2 \cup R_3) = w_a(R_1 \cup R_2 \cup R_3),$$
$$\alpha_2 w_a(R_2) + (1-\alpha_2) w_a(R_1 \cup R_3) = w_a(R_1 \cup R_2 \cup R_3).$$

Then

$$\alpha_1 w_a(R_1) + \alpha_2 w_a(R_2) + (1-\alpha_1-\alpha_2) w_a(R_3) = w_a(R_1 \cup R_2 \cup R_3).$$

To understand this, let us abbreviate $R_1 \cup R_2 \cup R_3$ as R_{123}, $R_1 \cup R_2$ as R_{12} and other terms similarly, and consider the following. By *Averaging*, $w_a(R_{123})$ is a weighted average of $w_a(R_1)$ and $w_a(R_{23})$, where the weight on $w_a(R_1)$ is $\rho_a^{R_1|R_{123}}$ and the weight on $w_a(R_{23})$ is $\rho_a^{R_{23}|R_{123}}$. Similarly, $w_a(R_{23})$ is a weighted average of $w_a(R_2)$ and $w_a(R_3)$, where the weight on $w_a(R_2)$ is $\rho_a^{R_2|R_{23}}$. Taking these two steps, it follows that $w_a(R_{123})$ is a weighted average of $w_a(R_1)$, $w_a(R_2)$ and $w_a(R_3)$, where the weights sum to one, the weight on $w_a(R_1)$ is $\rho^{R_1|R_{123}}$, the weight on $w_a(R_2)$ is $\rho^{R_2|R_{23}} \rho^{R_{23}|R_{123}}$, and the remaining weight is on $w_a(R_3)$. Following a similar procedure, but in a different order, first splitting $w_a(R_{123})$ into $w_a(R_2)$ and $w_a(R_{13})$, and then splitting $w_a(R_{13})$ into $w_a(R_1)$ and $w_a(R_3)$, we again arrive at $w_a(R_{123})$ as a weighted average of $w_a(R_1)$, $w_a(R_2)$ and $w_a(R_3)$. In this order of splitting the weights, the weight on $w_a(R_1)$ is $\rho^{R_1|R_{13}} \rho^{R_{13}|R_{123}}$. What *Order Independence* says is that these two ways of splitting $w_a(R_{123})$ lead to the same weighted average, so that in particular the weight on $w_a(R_1)$ is the same in both cases: $\rho^{R_1|R_{123}} = \rho^{R_1|R_{13}} \rho^{R_{13}|R_{123}}$. But this is just what *Chain Rule* says! Indeed, we have the following proposition.

Proposition 7. *Assume Simplification, No Information and Averaging. Then Chain Rule is equivalent to Order Independence.*

None of the assumptions that I have mentioned so far relate different decisions; each axiom provides a condition on w_a for all a, but none provides a constraint on the relation between w_a and w_b. The final assumption provides this relation.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Common Ratios. If $w_a(R) \neq 0$ and $w_b(R) \neq 0$,[58] then

$$\frac{w_a(R)}{w_a(\bar{R})} = \frac{w_b(R)}{w_b(\bar{R})}.$$

Recall from Section 5 that deliberative probabilities do not depend on the decision at hand – see equation (9). The role of *Common Ratios* is precisely to establish this independence of probabilities of reasons from actions.[59]

Let A_0 be the set of all actions *including* the default. A function $V : A_0 \times \Omega \to \mathbb{R}$ is a *value function*. I write $V_a(\omega)$ instead of $V(a, \omega)$; we can think of V_a as a function from Ω to $\mathbb{R}$ satisfying $V_a(\omega) = V(a, \omega)$. That is, for each a, V_a is a random variable. Given a probability measure P on Ω, I use the notation $\mathbb{E}_P$ to denote the expectation with respect to P. P is a *full support* probability measure if $P(\omega) > 0$ for all ω in Ω.[60] A reason R is *non-empty* if $R \neq \emptyset$.

Theorem 2. *Let* w *be a weight function and assume that* w *satisfies Simplification. Then the following conditions are equivalent.*

1. w *satisfies No Information, Averaging, Chain Rule, and Common Ratios.*
2. *There exists a value function* V *and a full support probability measure* P *on* Ω[61] *such that for all actions* a *in* A *and non-empty reasons* R

$$w_a(R) = \mathbb{E}_P[V_a - V_{\bar{a}} \,|\, R] - \mathbb{E}_P[V_a - V_{\bar{a}}]. \tag{17}$$

Moreover, if 1 holds, then in 2, P *is uniquely determined and* V *is unique up to a set of transformations that is explained in Proposition 8 of Appendix 1.*

The theorem shows that under our assumptions, weight can be taken as basic and deliberative probabilities and values derived. More precisely, (17) is a restatement of (3), which in Section 3.2, was used to define weight in terms of value and deliberative probability. So the theorem says that under the axioms, the weight function determines implicit values and probabilities such that weight, deliberative probability and value are related as in (3). We can take weight to be our fundamental object and think of values and deliberative probabilities as derived objects and still

[58] Our other assumptions imply that unless R is trivial, $w_a(R) \neq 0$. So the qualification 'If $w_a(R) \neq 0$ and $w_b(R) \neq 0$' is not necessary in the context of the other axioms. However, I include it because if we were to remove *Simplification*, then this qualification becomes important, and *Simplification* is an inessential assumption. I want to give the reader a sense for what the important assumptions are. Notice that *Simplification*, *No Information* and *Averaging* imply that for non-trivial reasons, the assumption that $w_a(R) \neq 0$ and $w_b(R) \neq 0$ implies that $w_a(\bar{R}) \neq 0$ and $w_b(\bar{R}) \neq 0$.

[59] Reasons may ultimately influence actions, but the independence of reasons and actions means that actions do not influence the deliberative probability of reasons.

[60] I write $P(\omega)$ instead of the more formally correct $P(\{\omega\})$.

[61] Formally, P is a probability measure on $(\Omega, 2^\Omega)$, where 2^Ω is the set of all subsets of Ω; recall that Ω is finite.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

derive a theory of weight that has the structure elaborated in the previous sections. That structure is characterized by the axioms. Conversely, we can, if we wish, take values and deliberative probabilities as basic. The theorem also shows that if we do so, then the resulting weight function will satisfy the axioms. Together the two directions of the theorem show that the axioms characterize the structure of the theory of weight studied in this paper.

Take the weight function as basic and impose the axioms so that implicit values and deliberative probabilities exist. It is instructive to consider how values and deliberative probabilities are derived from weight. Deliberative probabilities are derived from weight via equation (9) in Section 5.2. This relation between weight and deliberative probability was already discussed in that section. Proposition 8 in Appendix 1 shows that V in part 2 of the theorem can always be chosen so that $\mathbb{E}_P\left[\widehat{V}_a\right] = 0$, or in other words, so that neither a nor d has a default advantage over the other.[62] Then the comparative value $\widehat{V}_a(\omega)$ can expressed simply by[63]

$$\widehat{V}_a(\omega) = w_a(\{\omega\}). \tag{18}$$

This is somewhat artificial however, as it depends on Ω being finite and the family of reasons being the set of all subsets of Ω. So let us consider what would be the analogue of Theorem 2 if we allowed Ω to be an arbitrary set endowed with some sigma algebra $\mathscr{P}$ of reasons. To allow for the possibility of a non-atomic probability measure, we would have to eliminate *Simplification*. Then deliberative probabilities would still be derived from (9).[64] Let P be the deliberative probabilities derived from w in this case. Define $M_a(R) = w_a(R)P(R)$. Then M_a is a signed measure. Assume again that V is chosen so that $\mathbb{E}_P\left[\widehat{V}_a\right] = 0$. Then[65]

$$\widehat{V}_a = \frac{dM_a}{dP},$$

[62] Notice that the weight function w only determines how much a reason will move the decision. It does not by itself determine which alternative should ultimately by chosen. That also depends on the default advantages of the different actions in the deliberative position. See footnotes 25–26. So a complete theory of decision-making that took weights as fundamental would either have to assume initial neutrality or specify the default advantage δ_a of each action a (other than the default).

[63] More generally, when $\delta_a = \mathbb{E}_P\left[\widehat{V}_a\right]$ is not chosen to be zero – Proposition 8 in Appendix 1 shows that δ_a can be chosen freely: $\widehat{V}_a(\omega) = w_a(\{\omega\}) + \delta_a$.

[64] A complication would arise because it would not be possible sometimes that $w_a(R) = w_a(\bar{R}) = 0$. Deliberative probabilities would be derived from (9) for all reasons R for which $w_a(R) \neq 0$ and one would have to argue that one could extend P defined by (9) to be a probability measure defined on all R in $\mathscr{P}$, including those for which $w_a(R) = 0$.

[65] More generally, when $\delta_a = \mathbb{E}_P\left[\widehat{V}_a\right] \neq 0$, $\widehat{V}_a = \frac{dM_a}{dP} + \delta_a$.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

where the above is the Radon-Nikodym derivative of M_a with respect to P.

Let us reflect for a moment on the significance of the axioms and of Theorem 2 for justifying this paper's notion of weight. Regardless of whether one thinks that weight or value should be taken as fundamental, the axioms provide support for the particular notion of weight presented here. It seems that any plausible measure of weight should satisfy *No Information* and *Averaging*, which are satisfied by this paper's measure. The other axioms – *Chain Rule/Order Independence* and *Common Ratios* – provide a plausible way of filling out the picture, but they could be questioned. So the axioms also open the way to exploring other measures: One might want to search for interesting measures of weight that satisfy the former but not the latter axioms. It is important to note however that the measure presented here has desirable features: It handles overlap in a natural way, it allows for natural notions of conditioning and independence, it has a sharp and natural axiomatization,[66] and it provides a natural connection to decision theory. A measure of weight should not only be evaluated by its axioms but also by its consequences. Thus any competing notions should be evaluated by the yardstick of consequences as well. Moreover, it is my view that the deliberative position and the original definition of weight in terms of value and deliberative probability (Definition 2) are natural and appealing in themselves.

A final note: Above I have presented an axiomatization of weights. An alternative approach would be to start with a *strength* relation $\preccurlyeq_a$ on the set of potential reasons. $R_1 \preccurlyeq_a R_2$ would be read to say that R_2 is a stronger reason to take action a (as opposed to the default) than is reason R_1. One could then provide axioms on $\preccurlyeq_a$ from which weights – and hence also value and probabilities – can be derived.[67]

8. CONCLUSION: ATOMISM VS. HOLISM

In this concluding section I discuss the relevance of the model presented here for the issue of atomism vs holism about reasons. In the process, I also discuss some other related philosophical issues. I believe that this is

[66] There may be measures of weight that satisfy Averaging and No Information but violate the other axioms, which are difficult to axiomatize or have only unnatural or complicated axiomatizations.

[67] This is in the spirit of what Bolker (1967), Jeffrey (1990) and Domotor (1978) do for desirabilities. One could use their axioms as a starting point. One ingredient missing from the axiomatizations in these works that would need to be added would be an axiom to impose the *Common Ratios* property on weights of different actions. Notice finally that, in contrast to Theorem 2 for weight, an axiomatization of strength relations would not separately identify values and deliberative probabilities. In the case of desirability, Bolker (1967) showed that preferences over propositions do not separately identify desirabilities and probabilities. The mathematics is the same.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

just one of many philosophical debates to which the model presented here can contribute.

Dancy (2004) describes atomism as the thesis that 'a feature that is a reason in one case must remain a reason, and retain the same polarity, in any other', and holism as the negation of atomism. Atomism as stated is a very strong claim, and a more reasonable version would be something to the effect that what is a reason in one case will tend to be a reason and retain its polarity in other cases. Of course this weaker version of atomism is also vaguer. As the thesis of atomism is weakened in this way, holism, being the negation of atomism, is correspondingly strengthened. If we take atomism in the strict sense in which Dancy states it, then the current paper supports holism rather than atomism. For example, it can be the case that in the absence of other reasons, R is a reason for a ($w_a(R) > 0$), but given reasons $\mathscr{R}$, R is a reason against a ($w_a(R|\mathscr{R}) < 0$); and given reasons $\mathscr{S}$, R may again be a reason for a ($w_a(R|\mathscr{S}) > 0$). A further strengthening of atomism would be the claim that both the weight and polarity of a reason is the same in all contexts. This strengthening is a graded version of the original atomism thesis since it doesn't just make a claim that whether R is a reason – yes or no – is constant, but that the weight of the reason is constant. This graded version of atomism amounts to the additivity of weights, or equivalently the independence of reasons (see Proposition 3), and graded atomism is clearly false as a universal generalization in my model. In general, the magnitude of the weight of a reason R, $|w_a(R|\mathscr{R})|$, will depend on the class of reasons $\mathscr{R}$ that have already been taken into account. Theorem 1 characterizes the situations in which reasons are independent and their weights do not depend on other reasons.

Atomism in the strong sense stated by Dancy is a non-starter and I doubt that many people would sincerely attest to it upon reflection if it is not somehow qualified. (Dancy of course rejects atomism.) Without considering the model presented in this paper and the particular ways that I have chosen to model reasons, it is fairly obvious that it cannot be the case that the weight of all reasons is the same in all contexts. Independence of reasons cannot be a general thesis about reasons in the same way that probabilistic independence of events cannot be a general thesis about events. Rather, some events have probabilities that are not influenced by some other events and these events are probabilistically independent of one another. Similarly some reasons have weights that are not influenced by some other reasons and these reasons are independent of each other as regards their weight.

One might be tempted by the position that if we could just decompose complex reasons into basic components – which we might call atoms – then the weights and polarities of the atoms would always be independent of one another. Note that the term 'atom' is used in a different sense here

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

than in Dancy's definition. But perhaps this might give rise to a more reasonable version of atomism, namely that the weight and polarity of different *atomic* reasons are always independent of one another.[68] This is conceivable, but I am sceptical and I don't think that the model of this paper lends support to this form of atomism. Consider again the analogy to probability theory. In some cases – such as when we independently flip a fair coin many times – there may be a set of independent events out of which the probabilities of more complex events can be built. In general, however, probability measures can be very complex and probabilities of complex events may not admit a decomposition into probabilities of simple independent events. Similarly, in the case of reasons, there may be no useful way to decompose the set of all potential reasons into atomic reasons whose weights are constant across all configurations of other atomic reasons. The model of this paper does not particularly support a picture under which the weights of complex reasons are always built up in some way out of the weights of atomic reasons; it supports a picture rather under which there are constraints that link the weights of interrelated reasons, just like there are constraints that link the truth assignments to collections of logically interrelated propositions. In the same way that one might reject logical atomism, one might reject a version of reasons atomism that views the weights of all complex reasons as built out of the weights of independent atomic reasons. One would expect there to be a certain degree of decomposability of complex reasons into simple ones since otherwise it would be difficult for us to understand reasons, but it is difficult to say a priori what is the precise degree of such decomposabiltiy.

Fogal (2016) raises an issue that is distinct from the question of atomism, but that is related and bears on the question. Fogal points out that reasons may exhibit not just *inter*-contextual variability, but also *intra*-contextual variability. Fogal explains what he means by reference to an example from Schroeder (2007). Let R_1 be the proposition 'there will be dancing at the party' and R_2 be the proposition 'Billy loves to dance'. Fogal points out that it is acceptable to say that R_1 is a reason for Billy to go to the party and it is acceptable to say that R_2 is a reason for Billy to go to the party (and it is acceptable to say that the conjunction $R_1 \cap R_2$ is a reason to go to the party), but it is not acceptable to say that R_1 *is a reason for Billy to go to the party and so is* R_2. With regard to the unacceptable locution, Fogal writes 'The dominant reaction among those I've consulted is that such claims are guilty of something like 'double-counting''. Indeed, as discussed in the Introduction, the need to avoid double counting is one

[68] See Brown (2014) who studies atomism, which he defines as the view that 'all reasons supervene on atomic reasons' and related notions in a mereological framework. Brown contrasts atomism with 'wholism' as distinct from holism. Wholism is the view that all reasons reduce to overall reasons.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

of the main motivations for developing the model of this paper. Fogal develops an account in which reasons operate in 'normative clusters'. In the example, the normative cluster contains both R_1 and R_2. Either R_1 or R_2 can be taken as a *representative* of the cluster, which is why it is acceptable to say *individually* that R_1 is a reason or that R_2 is a reason. Fogal emphasizes that the part of the cluster that is not used as the representative serves as 'background information' and writes, 'whether a given fact can felicitously count as a reason in a particular context typically depends not only on there being other facts which taken together – and only taken together – fully explain the relevant normative status of an agent's action or attitude, but *also* on whether any of those other facts have already been cited as a reason.' The core of this account can be captured by my framework. The example that Fogal discusses is similar to the example that I discussed above in which the potential reasons are that I promised that I would go and that it is important to keep one's promises. In my discussion of the example in Section 6.3, I showed that in isolation, neither reason has any force ($w_a(R_1) = w_a(R_2) = 0$); it is only together as a 'cluster' that they have force. Conditioning on either reason – taking that reason as background – the other reason can be taken as the 'representative' of the cluster, and hence carry all of the weight of the cluster, so that as explained in Section 6.3, $w_a(R_1 \mid R_2) = w_a(R_2 \mid R_1) = w_a(R_1 \cap R_2) > 0$. In this way, my model can capture the intra-contextual variability that Fogal discusses.

Another way in which the model of this paper contributes to our understanding of the context-dependence of weights is that it provides a sense in which weights can be added even when reasons are not independent so that their weight depends on which other reasons have been taken into account. Regardless of whether reasons are independent, we can always add appropriate *conditional weights* (see Section 6.2). In general we cannot simply add unconditional weights, but we must interweave two operations: conditioning and addition. Except for the special case in which reasons are independent, when adding a weight, we must make sure that we have conditioned on all the reasons that we have already taken into account. That way, each weight measures the incremental change associated with taking a given reason into account given that other reasons have already been taken into account.

Dancy's motivation for discussing the issue of atomism vs holism is to settle another arguably more important question, namely: Are there any general moral principles? This is the debate between generalists and particularists. I believe that reflecting on the way complex interactions between weights of reasons are allowed in the model of this paper and the analogy to the interaction of propositions in probability theory can also contribute to the debate between generalists and particularists. However, that would be beyond the scope of the current paper.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

ACKNOWLEDGEMENTS

I am grateful to Dan Benjamin, Richard Bradley, Chris Chambers, Franz Dietrich, Caspar Hare, Christian List, Ben Lockwood, Barry Maguire, Anna Mahtani, Michael Mandler, Marcus Pivato, Tim Scanlon, Peter Sher, Shlomi Sher, Eran Shmaya, Justin Snedegar, Joel Sobel and participants at the Workshop on Reasons and Mental States in Decision Theory at the LSE, the second biennial Conference on Normative Ethics and Welfare Economics at Harvard Business School, and the inaugural PPE Society Meeting. I am also grateful to two anonymous referees.

REFERENCES

Alvarez, M. 2016. Reasons for action: justification, motivation, explanation. In *Stanford Encyclopedia of Philosophy* (Spring 2016 Edition. ed. E.N. Zalta. <http://plato.stanford.edu/archives/sum2016/entries/reasons-just-vs-expl/>.

Billingsley, P. 1995. *Probability and Measure*. New York, NY: John Wiley & Sons.

Bolker, E. D. 1966. Functions resembling quotients of measures. *Transactions of the American Mathematical Society* 124: 292–312.

Bolker, E. D. 1967. A simultaneous axiomatization of utility and subjective probability. *Philosophy of Science* 34: 333–340.

Broome, J. 2013. *Rationality through Reasoning*. Chichester: Wiley.

Brown, C. 2014. The composition of reasons. *Synthese* 191: 779–800.

Cariani, F. 2013. Aggregating with reason(s). *Synthese* 190: 1–25.

Crupi, V. and K. Tentori. 2016. Confirmation theory. In *The Oxford Handbook of Probability and Philosophy*, ed. A. Hájek and C. Hitchcock, 650–665. Oxford: Oxford University Press.

Dancy, J. 2004. *Ethics without Principles*. Oxford: Oxford University Press.

Dekel, E., B. L. Lipman and A. Rustichini. 1998. Standard state-space models preclude unawareness. *Econometrica* 66: 159–173.

Dietrich, F. and C. List. 2013. A reason-based theory of rational choice. *Noûs* 47: 104–134.

Dietrich, F. and List, C. 2016. Reason-based choice and context-dependence: an explanatory framework. *Economics and Philosophy* 32: 175–229.

Domotor, Z. 1978. Axiomatization of Jeffrey utilities. *Synthese* 39: 165–210.

Fogal, D. 2016. Reasons, reason, and context. In *Weighing Reasons*, ed. E. Lord and B. Maguire, 74–103. Oxford: Oxford University Press.

Halpern, J. Y. 2001. Alternative semantics for unawareness. *Games and Economic Behavior* 37: 321–339.

Harman, E. 2015. The irrelevance of moral uncertainty. *Oxford Studies in Metaethics* 10: 53–79.

Heifetz, A., M. Meier and B. C. Schipper. 2006. Interactive unawareness. *Journal of Economic Theory* 130: 78–94.

Horty, J. F. 2003. Reasoning with moral conicts. *Noûs* 37: 557–605.

Horty, J. F. 2007. Reasons as defaults. *Philosophers' Imprint* 7(3).

Horty, J. F. 2012. *Reasons as Defaults*. Oxford: Oxford University Press.

Jeffrey, R. C. 1990. *The Logic of Decision*. Chicago, IL: University of Chicago Press.

Kearns, S. and D. Star. 2009. Reasons as Evidence. *Oxford Studies in Metaethics* 4: 215–242.

Lockhart, T. 2000. *Moral Uncertainty and its Consequences*. New York, NY: Oxford University Press.

Lord, E. and B. Maguire, ed. 2016. *Weighing Reasons*. Oxford: Oxford University Press.

MacAskill, W. 2016. Normative uncertainty as a voting problem. *Mind* 125: 967–1004.

Maguire, B. 2016. The value-based theory of reasons. *Ergo* 3: 233–262.

Modica, S. and A. Rustichini. 1994. Awareness and partitional information structures. *Theory and Decision* 37: 107–124.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Nair, S. 2016. How do reasons accrue?, in *Weighing Reasons*, ed. E. Lord and B. Maguire, 56–73. Oxford: Oxford University Press.
Raz, J. 1999. *Practical Reason and Norms*. Oxford: Oxford University Press.
Scanlon, T. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
Schroeder, M. 2007. *Slaves of the Passions*. Oxford: Oxford University Press.
Sepielli, A. 2009. What to do when you don't know what to do. *Oxford Studies in Metaethics* 4: 5–28.
Snedegar, J. 2015. Contrastivism about reasons and ought. *Philosophy Compass* 10: 379–388.
Snedegar, J. 2018. Reasons for and reasons against. *Philosophical Studies* 175: 725–743.
Spohn, W. 2012. *The Laws of Belief: Ranking Theory and its Philosophical Applications*. Oxford: Oxford University Press.
Weatherson, B. 2014. Running risks morally. *Philosophical Studies* 167: 141–163.

APPENDIX 1. PROOFS OF THEOREMS

Proof of Theorem 1

Logical independence of $\mathscr{R}$ as I have defined it implies that for any truth assignment t and any subset J of $\{1, \ldots, n\}$, $P\left(\bigcap_{i\in J} R_i^{t(i)}\right) > 0$. I shall use this fact repeatedly in the proof without stating when I am using it.

First I prove that condition 2 in Theorem 1 implies condition 1. Choose any truth assignment t. Let $J \subseteq \{1, \ldots, n\}$ and $K = \{1, \ldots, n\} \backslash J$. Let T_K be the set of all truth assignments on K, or in other words, the set of all functions $s : K \to \{0, 1\}$. Then for any $i \in \{1, \ldots, n\}$,

$$\begin{aligned}
&\mathbb{E}\left[U_i \,\middle|\, \bigcap_{j\in J} R_j^{t(j)}\right] \\
&= \sum_{s\in T_K} \mathbb{E}\left[U_i \,\middle|\, \left(\bigcap_{j\in J} R_j^{t(j)}\right) \cap \left(\bigcap_{k\in K} R_k^{s(k)}\right)\right] P\left(\bigcap_{k\in K} R_k^{s(k)} \,\middle|\, \bigcap_{j\in J} R_j^{t(j)}\right) \\
&= \sum_{s\in T_K} \mathbb{E}\left[U_i \,\middle|\, \left(\bigcap_{j\in J} R_j^{t(j)}\right) \cap \left(\bigcap_{k\in K} R_k^{s(k)}\right)\right] P\left(\bigcap_{k\in K} R_k^{s(k)}\right),
\end{aligned}$$

where the second equality follows from condition 2a of the theorem. Now suppose that $i \in J$. Then

$$\begin{aligned}
&\sum_{s\in T_K} \mathbb{E}\left[U_i \,\middle|\, \left(\bigcap_{j\in J} R_j^{t(j)}\right) \cap \left(\bigcap_{k\in K} R_k^{s(k)}\right)\right] P\left(\bigcap_{k\in K} R_k^{s(k)}\right) \\
&= \sum_{s\in T_K} \mathbb{E}\left[U_i \,\middle|\, R_i^{t(i)}\right] P\left(\bigcap_{k\in K} R_k^{s(k)}\right) = \mathbb{E}\left[U_i \,\middle|\, R_i^{t(i)}\right],
\end{aligned}$$

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

where the second equality follows from condition 2(b)ii. On the other hand, if $i \notin J$, then

$$\begin{aligned}
&\sum_{s\in T_K} \mathbb{E}\left[U_i \,\middle|\, \left(\bigcap_{j\in J} R_j^{t(j)}\right) \cap \left(\bigcap_{k\in K} R_k^{s(k)}\right)\right] P\left(\bigcap_{k\in K} R_k^{s(k)}\right) \\
&\quad = \sum_{s\in T_K : s(i)=1} \mathbb{E}\left[U_i \,\middle|\, \left(\bigcap_{j\in J} R_j^{t(j)}\right) \cap \left(\bigcap_{k\in K} R_k^{s(k)}\right)\right] P\left(\bigcap_{k\in K} R_k^{s(k)}\right) \\
&\qquad + \sum_{s\in T_K : s(i)=0} \mathbb{E}\left[U_i \,\middle|\, \left(\bigcap_{j\in J} R_j^{t(j)}\right) \cap \left(\bigcap_{k\in K} R_k^{s(k)}\right)\right] P\left(\bigcap_{k\in K} R_k^{s(k)}\right) \\
&\quad = \sum_{s\in T_K : s(i)=1} \mathbb{E}\left[U_i \,|\, R_i\right] P\left(\bigcap_{k\in K} R_k^{s(k)}\right) + \sum_{s\in T_K : s(i)=0} \mathbb{E}\left[U_i \,|\, \bar{R}_i\right] P\left(\bigcap_{k\in K} R_k^{s(k)}\right) \\
&\quad = \mathbb{E}\left[U_i \,|\, R_i\right] P\left(R_i\right) + \mathbb{E}\left[U_i \,|\, \bar{R}_i\right] P\left(\bar{R}_i\right) = \mathbb{E}\left[U_i\right],
\end{aligned}$$

where the second equality follows from 2(b)ii. Putting this all together,

$$\mathbb{E}\left[U_i \,\middle|\, \bigcap_{j\in J} R_j^{t(j)}\right] = \begin{cases} \mathbb{E}\left[U_i \,\middle|\, R_i^{t(i)}\right], & \text{if } i \in J, \\ \mathbb{E}\left[U_i\right], & \text{if } i \notin J. \end{cases} \tag{19}$$

Now let $\{i_1, \ldots, i_n\}$ be a permutation of $\{1, \ldots, n\}$, and let $k \in \{2, \ldots, n\}$. We have

$$\begin{aligned}
&\mathbb{E}\left[\widehat{V}_a \,\middle|\, \bigcap_{j=1}^{k} R_{i_j}^{t(i_j)}\right] - \mathbb{E}\left[\widehat{V}_a \,\middle|\, \bigcap_{j=1}^{k-1} R^{t(i_j)}\right] \\
&\quad = \sum_{\ell=1}^{n} \mathbb{E}\left[U_{i_\ell} \,\middle|\, \bigcap_{j=1}^{k} R_{i_j}^{t(i_j)}\right] - \sum_{\ell=1}^{n} \mathbb{E}\left[U_{i_\ell} \,\middle|\, \bigcap_{j=1}^{k-1} R^{t(i_j)}\right] \\
&\quad = \left(\sum_{\ell=1}^{k} \mathbb{E}\left[U_{i_\ell} \,\middle|\, R_{i_\ell}^{t(i_\ell)}\right] + \sum_{\ell=k+1}^{n} \mathbb{E}\left[U_{i_\ell}\right]\right) - \left(\sum_{\ell=1}^{k-1} \mathbb{E}\left[U_{i_\ell} \,\middle|\, R_{i_\ell}^{t(i_\ell)}\right] + \sum_{\ell=k}^{n} \mathbb{E}\left[U_{i_\ell}\right]\right) \qquad (20) \\
&\quad = \mathbb{E}\left[U_{i_k} \,\middle|\, R_{i_k}^{t(i_k)}\right] - \mathbb{E}\left[U_{i_k}\right] \\
&\quad = \mathbb{E}\left[U_{i_k} \,\middle|\, R_{i_k}^{t(i_k)}\right] - \mathbb{E}\left[U_{i_k}\right] + \sum_{i_j \neq i_k} \left(\mathbb{E}\left[U_{i_j} \,\middle|\, R_{i_k}^{t(i_k)}\right] - \mathbb{E}\left[U_{i_j}\right]\right) \\
&\quad = \sum_{j=1}^{n} \left(\mathbb{E}\left[U_{i_j} \,\middle|\, R_{i_k}^{t(i_k)}\right] - \mathbb{E}\left[U_{i_j}\right]\right) = \mathbb{E}\left[\widehat{V}_a \,\middle|\, R_{i_k}^{t(i_k)}\right] - \mathbb{E}\left[\widehat{V}_a\right],
\end{aligned}$$

where the first and last equalities follow from condition 2(b)i, and the second and fourth equalities follow from (19) – in the case of the fourth equality by choosing $J = \{i_k\}$. However (20) amounts to strong *a*-independence. So this completes the proof that condition 2 in Theorem 1 implies condition 1.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Next I prove that condition 1 of Theorem 1 implies condition 2. So assume that $\mathscr{R}$ is strongly a-independent. I start by proving that $R_1, \ldots, R_n$ are mutually probabilistically independent. To do so, it is sufficient to show that for all $J \subseteq \{1, \ldots, n\}$

$$P\left(\bigcap_{i\in J} R_i\right) = \prod_{i\in J} P\left(R_i\right). \tag{21}$$

I prove (21) by induction on $k = |\mathscr{S}|$. The conclusion is immediate when $k = 1$. So suppose that it is true for $k-1$, and I prove it for k. I consider the case in which $\mathscr{S} = \{R_1, \ldots, R_k\}$, but the proof will obviously extend to any $\mathscr{S}$ with $|\mathscr{S}| = k$.

Because $\mathscr{R}$ does not contain a-empty reasons, there exists $J \subseteq \{1, \ldots, n\} \backslash k$ such that $\mathbb{E}\left[\widehat{V}_a \,\middle|\, R_k \cap \left(\bigcap_{j\in J} R_j\right)\right] \neq \mathbb{E}\left[\widehat{V}_a \,\middle|\, \bigcap_{j\in J} R_j\right]$. Strong a-independence implies that $\mathbb{E}\left[\widehat{V}_a \,|R_k\right] \neq \mathbb{E}\left[\widehat{V}_a\right]$. By the law of iterated expectations:

$$\mathbb{E}\left[\widehat{V}_a\right] = \mathbb{E}\left[\widehat{V}_a \,|R_k\right] P\left(R_k\right) + \mathbb{E}\left[\widehat{V}_a \,|\bar{R}_k\right] P\left(\bar{R}_k\right). \tag{22}$$

It follows that

$$\mathbb{E}\left[\widehat{V}_a \,|R_k\right] \neq \mathbb{E}\left[\widehat{V}_a \,|\bar{R}_k\right]. \tag{23}$$

Define $R^{k-1} := \bigcap_{j=1}^{k-1} R_j$. Strong a-independence implies

$$\mathbb{E}\left[\widehat{V}_a \,\middle|\, R_k \cap R^{k-1}\right] - \mathbb{E}\left[\widehat{V}_a \,\middle|\, R^{k-1}\right] = \mathbb{E}[\widehat{V}_a | R_k] - \mathbb{E}[\widehat{V}_a]. \tag{24}$$

Plugging (22) into (24) and rearranging terms,

$$\begin{aligned}\mathbb{E}\left[\widehat{V}_a \,\middle|\, R_k \cap R^{k-1}\right] &= \mathbb{E}\left[\widehat{V}_a \,|R_k\right] + \mathbb{E}\left[\widehat{V}_a \,\middle|\, R^{k-1}\right] \\ &\quad - \left(\mathbb{E}\left[\widehat{V}_a \,|R_k\right] P\left(R_k\right) + \mathbb{E}\left[\widehat{V}_a \,|\bar{R}_k\right] P\left(\bar{R}_k\right)\right) \\ &= \mathbb{E}\left[\widehat{V}_a \,\middle|\, R^{k-1}\right] + \left(\mathbb{E}\left[\widehat{V}_a \,|R_k\right] - \mathbb{E}\left[\widehat{V}_a \,|\bar{R}_k\right]\right) P\left(\bar{R}_k\right).\end{aligned} \tag{25}$$

Similarly,

$$\mathbb{E}\left[\widehat{V}_a \,\middle|\, \bar{R}_k \cap R^{k-1}\right] = \mathbb{E}\left[\widehat{V}_a \,\middle|\, R^{k-1}\right] + \left(\mathbb{E}\left[\widehat{V}_a \,|\bar{R}_k\right] - \mathbb{E}\left[\widehat{V}_a \,|R_k\right]\right) P\left(R_k\right). \tag{26}$$

Again, by the law of iterated expectations,

$$\begin{aligned}\mathbb{E}\left[\widehat{V}_a \,\middle|\, R^{k-1}\right] =& \mathbb{E}\left[\widehat{V}_a \,\middle|\, R_k \cap R^{k-1}\right] P\left(R_k \,\middle|\, R^{k-1}\right) \\ &+ \mathbb{E}\left[\widehat{V}_a \,\middle|\, \bar{R}_k \cap R^{k-1}\right] P\left(\bar{R}_k \,\middle|\, R^{k-1}\right).\end{aligned} \tag{27}$$

Plugging (25) and (26) into (27), we get:

$$\begin{aligned}\mathbb{E}\left[\widehat{V}_a \,\middle|\, R^{k-1}\right] = &\left[\mathbb{E}\left[\widehat{V}_a \,\middle|\, R^{k-1}\right] + \left(\mathbb{E}\left[\widehat{V}_a \,|R_k\right] - \mathbb{E}\left[\widehat{V}_a \,|\bar{R}_k\right]\right) P\left(\bar{R}_k\right)\right] P\left(R_k \,\middle|\, R^{k-1}\right) \\ &+ \left[\mathbb{E}\left[\widehat{V}_a \,\middle|\, R^{k-1}\right] + \left(\mathbb{E}\left[\widehat{V}_a \,|\bar{R}_k\right] - \mathbb{E}\left[\widehat{V}_a \,|R_k\right]\right) P\left(R_k\right)\right] P\left(\bar{R}_k \,\middle|\, R^{k-1}\right).\end{aligned}$$

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

This is equivalent to

$$\left(\mathbb{E}\left[\widehat{V}_a \,|R_k\right] - \mathbb{E}\left[\widehat{V}_a \,|\bar{R}_k\right]\right) P\left(\bar{R}_k\right) P\left(R_k \,|R^{k-1}\right)$$
$$= \left(\mathbb{E}\left[\widehat{V}_a \,|R_k\right] - \mathbb{E}\left[\widehat{V}_a \,|\bar{R}_k\right]\right) P\left(R_k\right) P\left(\bar{R}_k \,|R^{k-1}\right),$$

which, by (23), is equivalent to $P\left(\bar{R}_k\right) P\left(R_k \,|R^{k-1}\right) = P\left(R_k\right) P\left(\bar{R}_k \,|R^{k-1}\right)$, which in turn is equivalent to $P\left(R^{k-1} \,|R_k\right) = P\left(R^{k-1} \,|\bar{R}_k\right)$. By the law of total probability:

$$P\left(R^{k-1}\right) = P\left(R^{k-1} \,|R_k\right) P\left(R_k\right) + P\left(R^{k-1} \,|\bar{R}_k\right) P\left(\bar{R}_k\right).$$

It now follows that $P(R^{k-1} \,|\, R_k) = P(R^{k-1})$, or equivalently $P(R_k \cap R^{k-1}) = P(R_k)P(R^{k-1})$, but note that

$$P\left(R_k \cap R^{k-1}\right) = P\left(R_k\right) P\left(R^{k-1}\right) \Leftrightarrow P\left(\bigcap_{j=1}^{k} R_j\right) = P\left(R_k\right) P\left(\bigcap_{j=1}^{k-1} R_j\right)$$
$$\Leftrightarrow P\left(\bigcap_{j=1}^{k} R_j\right) = \prod_{j=1}^{k} P(R_j),$$

where the last line uses the inductive hypothesis. This establishes (21) when $\mathscr{S} = \{R_1, \ldots, R_k\}$. As the proof is similar whenever $|\mathscr{S}| = k$, we have established by induction that $R_1, \ldots, R_n$ are mutually probabilistically independent.

For $i = 1, \ldots n$, define $U_i := w_a\left(R_i\right) 1_{R_i} + w_a\left(\bar{R}_i\right) 1_{\bar{R}_i}$ for $i = 1, \ldots, n-1$, where for any event R, 1_R is the indicator function for R, and define $U_n := \widehat{V}_a - \sum_{j=1}^{n-1} U_j$. Then, it is immediate that $\widehat{V}_a = \sum_{i=1}^{n} U_i$. Moreover, it is immediate that condition 2(b)ii of the theorem holds for U_i when $i = 1, \ldots, n-1$. We have only to prove 2(b)ii when $i = n$. Before proceeding, it is useful to observe that by the law of iterated expectations,

$$w_a\left(R\right) P\left(R\right) + w_a\left(\bar{R}\right) P\left(\bar{R}\right) = 0. \tag{28}$$

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Now choose a truth assignment t. We have:

$$\begin{aligned}
\mathbb{E}\left[U_n \middle| \bigcap_{i=1}^{n} R_i^{t(i)}\right] &= \mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{n} R_i^{t(i)}\right] - \sum_{j=1}^{n-1} \mathbb{E}\left[U_j \middle| \bigcap_{i=1}^{n} R_i^{t(i)}\right] \\
&= \mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{n} R_i^{t(i)}\right] - \sum_{j=1}^{n-1} \mathbb{E}\left[w_a\left(R_j\right) 1_{R_j} + w_a\left(\bar{R}_j\right) 1_{\bar{R}_j} \middle| \bigcap_{i=1}^{n} R_i^{t(i)}\right] \\
&= \mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{n} R_i^{t(i)}\right] - \sum_{j=1}^{n-1} w_a\left(R_j^{t(j)}\right) \\
&= \mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{n} R_i^{t(i)}\right] - \sum_{j=1}^{n-1} \left(\mathbb{E}\left[\widehat{V}_a \middle| R_j^{t(j)}\right] - \mathbb{E}\left[\widehat{V}_a\right]\right) \\
&= \mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{n} R_i^{t(i)}\right] - \sum_{j=1}^{n-1} \left(\mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{j} R_i^{t(i)}\right] - \mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{j-1} R_i^{t(i)}\right]\right) \\
&= \left(\mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{n} R_i^{t(i)}\right] - \mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{n-1} R_i^{t(i)}\right]\right) \\
&\quad + \sum_{j=1}^{n-2} \left(\mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{j} R_i^{t(i)}\right] - \mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{j} R_i^{t(i)}\right]\right) + \mathbb{E}\left[\widehat{V}_a\right] \\
&= \left(\mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{n} R_i^{t(i)}\right] - \mathbb{E}\left[\widehat{V}_a \middle| \bigcap_{i=1}^{n-1} R_i^{t(i)}\right]\right) + \mathbb{E}\left[\widehat{V}_a\right] \\
&= \left(\mathbb{E}\left[\widehat{V}_a \middle| R_n^{t(n)}\right] - \mathbb{E}\left[\widehat{V}_a\right]\right) + \mathbb{E}\left[\widehat{V}_a\right] = \mathbb{E}\left[\widehat{V}_a \middle| R_n^{t(n)}\right] \\
&= \mathbb{E}\left[\widehat{V}_a \middle| R_n^{t(n)}\right] - \sum_{i=1}^{n-1} \left[w_a\left(R_i\right) P\left(R_i\right) + w_a\left(\bar{R}_i\right) P\left(\bar{R}_i\right)\right] \\
&= \mathbb{E}\left[\widehat{V}_a \middle| R_n^{t(n)}\right] - \sum_{i=1}^{n-1} \mathbb{E}\left[w_a\left(R_i\right) 1_{R_i} + w_a\left(\bar{R}_i\right) 1_{\bar{R}_i}\right] \\
&= \mathbb{E}\left[\widehat{V}_a \middle| R_n^{t(n)}\right] - \sum_{i=1}^{n-1} \mathbb{E}\left[w_a\left(R_i\right) 1_{R_i} + w_a\left(\bar{R}_i\right) 1_{\bar{R}_i} \middle| R_n^{t(n)}\right] \\
&= \mathbb{E}\left[\widehat{V}_a \middle| R_n^{t(n)}\right] - \sum_{i=1}^{n-1} \mathbb{E}\left[U_i \middle| R_n^{t(n)}\right] = \mathbb{E}\left[\widehat{V}_a - \sum_{i=1}^{n-1} U_i \middle| R_n^{t(n)}\right] = \mathbb{E}\left[U_n \middle| R_n^{t(n)}\right].
\end{aligned}$$

In the above derivation, the fifth equality follows from strong a-independence, the tenth equality follows from (28), and the twelfth equality follows from the mutual probabilistic independence of $R_1, \ldots, R_n$, which was established above. This derivation establishes that U_n satisfies condition 2(b)ii, and completes the proof that condition 1 of Theorem 1 implies condition 2. □

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Proof of Theorem 2

Before proceeding to the proof, I state the uniqueness properties of V.

Proposition 8. *If in Theorem 2, statement 1 holds, then in statement 2, and for any collection of numbers* $(\delta_{\mathrm{a}} : \mathrm{a} \in \mathrm{A})$ *it is consistent to impose the condition* $\mathbb{E}[V_a - V_d] = \delta_a, \forall a \in A$, *and once this is done, for all* $\mathrm{a} \in \mathrm{A}$, $\widehat{V}_a = V_a - V_d$ *is uniquely determined, while* $\mathrm{V_d}$ *can be chosen freely. Thus by setting* $\delta_{\mathrm{a}} = 0$ *for all* $\mathrm{a} \in \mathrm{A}$, *we may always impose initial neutrality consistently with* w.

In Proposition 8, we can interpret the δ_as as default advantages. That is, for any action a, δ_a is the advantage that a starts at relative to the default if no reasons are taken into account.

I now proceed to prove Theorem 2 as well as Proposition 8. It is straightforward to verify that part 2 of the theorem implies part 1. Therefore I omit this direction and prove only that part 1 implies part 2. That is, I show that if the weight w function obeys the axioms, then there exists full support probability measure P and value function V that satisfies (17).

As discussed in Section 5.2, deliberative probability P can be defined in terms of weight w via (9). *Common Ratios* implies $P(R)$ as defined in (9) does not depend on the choice of action a. That is, if w_a is replaced by w_b in (9), where b is some other action, the value of $P(R)$ does not change. *Simplification* implies that the right-hand side of (9) is well defined for all R.[69] We need to prove that P is indeed a probability measure. Observe first that (9) implies that $P(R) \geqslant 0$ for all R, and that (9), *No Information*, and *Simplification* imply that $P(\Omega) = 1$. For full support, observe that *Simplification*, *No Information*, *Averaging*, and (9) imply that $P(\omega) > 0$ for all ω. To prove that P is a probability measure, we now need only show that P is additive. *Simplification*, *No Information*, and *Averaging* imply that $w_a(\Omega) = 0$, $w_a(\varnothing) \neq 0$, and for all non-trivial R, $w_a(R) > 0 \Leftrightarrow w_a\left(\bar{R}\right) < 0$. It follows that $\left|\frac{w_a(R)}{w_a\left(\bar{R}\right)}\right| = -\frac{w_a(R)}{w_a\left(\bar{R}\right)}$. Using this fact and multiplying the numerator and denominator of the right-hand side of (9) by $w_a\left(\bar{R}\right)$ (when $R \neq \varnothing$), we get $P(R) = \frac{w_a\left(\bar{R}\right)}{w_a\left(\bar{R}\right) - w_a(R)}$, for all non-empty R. Also, $P(\emptyset) = 0 = \frac{w_a(\Omega)}{w_a(\Omega) - w_a(\emptyset)}$.[70] As was mentioned in Section 7, it follows from (16), the definition of $\rho^{R|S}$ (for $R \subseteq S \subseteq \Omega$ with $S \neq \varnothing$), that $\rho_a^{R|S} = \frac{w_a(S) - w_a(S-R)}{w_a(R) - w_a(S-R)}$. It now follows from *No Information* that

$$P(R) = \rho_a^{R|\Omega}. \tag{29}$$

[69] Suppose that $w_a\left(\bar{R}\right) = 0$. Then by *Simplification*, $w_a(R) \neq 0$. So $\frac{w_a(R)}{w_a\left(\bar{R}\right)} = \infty$ by the convention that for all real numbers r with $r \neq 0$, $\frac{r}{0} = \infty$. So $P(R) = \frac{1}{1+\infty} = 0$, again by convention. Incidentally, *Simplification*, *No Information* and *Averaging* imply that $w_a\left(\bar{R}\right) = 0$ only if $R = \varnothing$.

[70] To see why $P(\varnothing) = 0$, consult footnote 69.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

At this point, it is useful to observe that if $R_1 \cap R_2 = \varnothing$ and $R_1 \cup R_2 \neq \varnothing$,

$$\begin{aligned}\rho_a^{R_1|R_1\cup R_2} + \rho_a^{R_2|R_1\cup R_2} &= \frac{w_a(R_1 \cup R_2) - w_a(R_2)}{w_a(R_1) - w_a(R_2)} + \frac{w_a(R_1 \cup R_2) - w_a(R_1)}{w_a(R_2) - w_a(R_1)} \\ &= \frac{w_a(R_1 \cup R_2) - w_a(R_2)}{w_a(R_1) - w_a(R_2)} + \frac{w_a(R_1) - w_a(R_1 \cup R_2)}{w_a(R_1) - w(R_2)} = 1.\end{aligned} \tag{30}$$

Next observe that if $R_1 \cap R_2 = \varnothing$,[71]

$$\begin{aligned}P(R_1) + P(R_2) &= \rho_a^{R_1|\Omega} + \rho_a^{R_2|\Omega} = \rho_a^{R_1|R_1\cup R_2}\rho_a^{R_1\cup R_2|\Omega} + \rho_a^{R_2|R_1\cup R_2}\rho_a^{R_1\cup R_2|\Omega} \\ &= \left(\rho_a^{R_1|R_1\cup R_2} + \rho_a^{R_2|R_1\cup R_2}\right)\rho_a^{R_1\cup R_2|\Omega} = \rho_a^{R_1\cup R_2|\Omega} = P(R_1 \cup R_2),\end{aligned} \tag{31}$$

where the first and last equality follow from (29), the second equality follows from *Chain Rule*, and the fourth equality uses (30). This establishes that P is additive and hence a probability measure.

Next we must construct V for which (17). On the way to this goal, consider R_1 and R_2 with $R_1 \cap R_2 = \varnothing$, and observe that[72]

$$\begin{aligned}w_a(R_1)P(R_1) + w_a(R_2)P(R_2) &= w_a(R_1)\rho_a^{R_1|\Omega} + w_a(R_2)\rho_a^{R_2|\Omega} \\ &= \rho_a^{R_1|R_1\cup R_2}\rho_a^{R_1\cup R_2|\Omega}w_a(R_1) + \rho_a^{R_2|R_1\cup R_2}\rho_a^{R_1\cup R_2|\Omega}w_a(R_2) \\ &= \left(\rho_a^{R_1|R_1\cup R_2}w_a(R_1) + \rho_a^{R_2|R_1\cup R_2}w_a(R_2)\right)\rho_a^{R_1\cup R_2|\Omega} \\ &= w_a(R_1 \cup R_2)\rho_a^{R_1\cup R_2|\Omega} = w_a(R_1 \cup R_2)P(R_1 \cup R_2),\end{aligned} \tag{32}$$

where the first and last equality follow from (29), the second equality follows from *Chain Rule*, and the fourth equality follows from (16). So we have shown that

$$\begin{aligned}&\forall R_1, R_2 \subseteq \Omega, \text{ if } R_1 \cap R_2 = \emptyset, \\ &\quad \text{then } w_a(R_1)P(R_1) + w_a(R_2)P(R_2) = w_a(R_1 \cup R_2)P(R_1 \cup R_2).\end{aligned} \tag{33}$$

Next choose an arbitrary function $V_d : \Omega \to \mathbb{R}$ and a vector $(\delta_a : a \in A)$. Define

$$V_a(\omega) = w_a(\{\omega\}) + \delta_a + V_d(\omega). \tag{34}$$

[71] As the derivation in (31) relies on (30) and (30) assumes that $R_1 \cup R_2 \neq 0$, to derive additivity of P, we must separately consider the case where $R_1 \cup R_2 = \varnothing$, or equivalently the case in which $R_1 = R_2 = \varnothing$. However in this case, additivity follows immediately from the fact that $P(\varnothing) = 0$.

[72] The derivation in (32) holds when $R_1 \cup R_2 \neq \varnothing$. If $R_1 = R_2 = \varnothing$, then $P(R_1) = P(R_2) = P(R_1 \cup R_2) = 0$ – see footnote 71 – so that in this case too $w_a(R_1)P(R_1) + w_a(R_2)P(R_2) = w_a(R_1 \cup R_2)P(R_1 \cup R_2)$.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

Then choose non-empty $R \subseteq \Omega$. We have

$$\begin{aligned}
w_a(R) &= \sum_{\omega \in R} w_a(\{\omega\}) \frac{P(\{\omega\})}{P(R)} = \sum_{\omega \in R} (V_a(\omega) - V_d(\omega) - \delta_a) \frac{P(\{\omega\})}{P(R)} \\
&= \sum_{\omega \in R} (V_a(\omega) - V_d(\omega)) \frac{P(\{\omega\})}{P(R)} - \delta_a \\
&= \sum_{\omega \in R} (V_a(\omega) - V_d(\omega)) \frac{P(\{\omega\})}{P(R)} - \sum_{\omega \in \Omega} \delta_a P(\{\omega\}) \\
&= \sum_{\omega \in R} (V_a(\omega) - V_d(\omega)) \frac{P(\{\omega\})}{P(R)} - \sum_{\omega \in \Omega} (V_a(\omega) - V_d(\omega) - w_a(\{\omega\})) P(\{\omega\}) \\
&= \sum_{\omega \in R} (V_a(\omega) - V_d(\omega)) \frac{P(\{\omega\})}{P(R)} - \sum_{\omega \in \Omega} (V_a(\omega) - V_d(\omega)) P(\{\omega\}) \\
&\quad - \sum_{\omega \in \Omega} w_a(\{\omega\}) P(\{\omega\}) \\
&= \sum_{\omega \in R} (V_a(\omega) - V_d(\omega)) \frac{P(\{\omega\})}{P(R)} - \sum_{\omega \in \Omega} (V_a(\omega) - V_d(\omega)) P(\{\omega\}) \\
&\quad - w_a(\Omega) P(\Omega) \\
&= \mathbb{E}_P[V_a - V_d \,|R] - \mathbb{E}_P[V_a - V_d],
\end{aligned} \tag{35}$$

where the first and seventh equalities follow from repeated applications of (33), the second and fifth equalities follow from (34), and $w_a(\Omega)P(\Omega)$ disappears in the last equality by *No Information*. This establishes (17).

Next I establish the uniqueness of P and Proposition 8. That P must satisfy (9) follows from the law of iterated expectations.[73] Observe next that if V_a is chosen as in (34), then *No Information* and (33) imply that $\mathbb{E}_P\left[\widehat{V}_a\right] = \delta_a$. Since the derivation (35) applies for an arbitrary choice of δ_a and V_d, it follows that δ_a and V_d can be chosen arbitrarily. However, once this is done, $V_a(\omega)$ is pinned down by (34), and any change in $V_a(\omega)$ would lead to $w_a(\{\omega\}) \neq \mathbb{E}_P\left[\widehat{V}_a \,|\{\omega\}\right] - \mathbb{E}_P\left[\widehat{V}_a\right]$. This completes the proof. □

APPENDIX 2. SOME RELATIONS TO THE LITERATURE

In this appendix, I describe in more detail some relations between my work and other formal work pertaining to reasons. I also discuss the relation of deliberative probability to the literature on normative uncertainty.

First I briefly contrast my model with that of Dietrich and List (2013). In Dietrich and List's model, a reason R is a proposition that is true of some

[73] See footnote 31 and recall that *Simplification*, *No Information* and *Averaging* imply that for all non-trivial R, $w_a(R) \neq 0$.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

alternatives. A reason can be thought of as an attribute or property of those alternatives. Thus, it can be modelled as a set of alternatives. In contrast, in my model, a reason is a set of states. Consider in particular the proposition R that I promised Jim that I would go to the movies. R is a reason to go to the movies and it is also a reason to apologize if I don't go. Thus, the proposition R can be a reason for different actions, going and apologizing, and it relates to these different actions in different ways. This suggests that reasons may not always be best thought of as attributes of alternatives. This is one reason that I have chosen to model reasons as propositions that are not necessarily about the alternatives. Dietrich and List define a preference relation $\succsim_{\mathscr{R}}$ over alternatives where $\mathscr{R}$ is the set of reasons that are motivating (under a descriptive interpretation) or the set of reasons taken to be normatively compelling (under a normative interpretation). They also define another relation $\geqslant$ directly over sets of reasons themselves. Deitrich and List provide axioms under which the relation $\succsim_{\mathscr{R}}$ over alternatives is determined by the relation $\geqslant$ over bundles of reasons in $\mathscr{R}$. These axioms have to do with how $\succsim_{\mathscr{R}}$ changes as $\mathscr{R}$ changes. As can be seen from this description, Dietrich and List's model is very different from the model I have presented in this paper. To take just one example, the various properties on weights that I specify in Section 7 need not hold in Dietrich and List's model.

I have already discussed the relationship to Bolker-Jeffrey decision theory in Section 7. I now discuss the connection to confirmation theory. As mentioned in the Introduction, both Spohn (2012) and Nair (2016) suggest quantitative and qualitative relations between a proposition B confirming a proposition A and B's being a reason to believe A. Confirmation is usually defined probabilistically, and in addition to relating the standard probabilistic definition of confirmation to being a reason, Spohn develops an analogue of this notion in ranking theory, his preferred epistemic model. Nair writes, 'The connection [to reasons] can be made precise by noting that, plausibly, p is a reason to believe q just in case p confirms q. This is especially plausible once we bear in mind that accepting this connection does not force us to accept that reasons can be analysed in terms of confirmation.' According to the incremental notion of confirmation, evidence E confirms hypothesis H if $P(H \mid E) > P(H)$.

Spohn refers to $P(A)$ and $P(\bar{A})$ as *weights* and $P(B \mid A) - P(B)$ and $P(B \mid \bar{A}) - P(B)$ as *strengths*, which combine to make a *beam balance* equation (the law of total probability). Nair writes that we can get quantitative notions of weight by appeal to confirmation measures. Nair (2016) is specifically concerned with the relation to the 'accrual' of the weight of reasons and the question of when weights can be added. Nair proposes that the independence of reasons has to do with whether reasons are basic or derivative. I have expressed scepticism about this approach in Section 8, although I believe it would be fruitful to attempt to develop it.

Three of the main confirmation measures are the *probability difference* $P(H \mid E) - P(H)$, the *probability ratio* $\frac{P(H|E)}{P(H)}$ and the *likelihood ratio* $\frac{P(E|H)}{P(E|\bar{H})}$ (see Crupi and Tentori 2016).[74] To make the connection to my model precise, consider the special

[74] There are many other confirmation measures that are ordinally equivalent to one of these in the sense that they rank confirmation in hypothesis evidence pairs (H, E) in the same way.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

case in which $V_a = 1_A$ where 1_A is the indicator function that takes value 1 in states ω in A and 0 in states ω outside of A. The interpretation is that a leads to a good outcome if A is true and a bad outcome if A is false. Suppose also that $V_d \equiv 0$. Then

$$w_a(R) = \mathbb{E}\left[\widehat{V}_a \,|R\right] - \mathbb{E}\left[\widehat{V}_a\right] = \mathbb{E}[1_A|R] - \mathbb{E}[1_A] = P(A|R) - P(A).$$

So in this special case, R is a reason to select a if and only if R confirms A and the weight of R is the probability difference measure of confirmation.[75] However, note that the interpretation here is not that R is a reason to believe A, but rather that R is a reason to do a.

I now discuss the relation of deliberative uncertainty, and specifically deliberative uncertainty of normative propositions – see Section 5.1 – to the literature on normative uncertainty. Some references to that literature can be found in footnote 29. One important respect in which deliberative uncertainty about normative propositions in this paper differs from the normative uncertainty studied in this literature is that here the agent can enter a state of deliberative uncertainty when she knows which moral propositions are true. The agent compares how she should evaluate her decision given that she knows a normative proposition to how she should evaluate the decision if she did not know this – and in the case of highly significant propositions to her evaluation conditional on assuming that the proposition is very unlikely, which is close to assuming that it is false. Some of the issues and controversies associated with normative uncertainty may arise also for deliberative uncertainty about normative propositions. Some authors believe that the right thing to do under normative uncertainty is to maximize expected value subject to that uncertainty and others believe that the right thing to do is whatever the true moral theory says is better or right. There is also the problem of intertheoretic comparability of value, analogous to the classic problem of interpersonal comparability of utility. I do not have space to explore these issues here as they pertain to deliberative probability. I restrict myself to two comments. First, it would be problematic to say that one cannot reason hypothetically on the supposition that a certain moral principle is false, and it is intuitive that if one thinks that a moral principle is very unlikely, one should come to very similar conclusions to those one would come to if one thought that the moral principle were false. This opens up the possibility of reasoning hypothetically under normative uncertainty more generally. There are objections to the sort of continuity I am positing here, but I will have to leave those aside. Second, probabilities and values need not be taken as primitive. It is possible to start from weights, and then probabilities and values are implicitly defined (see Section 7). So if one grants that it is possible to assign weights to normative reasons, then one can understand values conditional on those reasons and deliberative probabilities of those reasons in terms of those weights.

[75] Note however that this is the probability difference measure for a *fixed* hypothesis.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press

BIOGRAPHICAL INFORMATION

Itai Sher is an Associate Professor of Economics at the University of Massachusetts Amherst. His research focuses on normative issues at the boundary of ethics and economics. He is a founder and co-organizer of the conference series 'Normative Ethics and Welfare Economics'.

https://doi.org/10.1017/S0266267118000160 Published online by Cambridge University Press