Comparative Value and the Weight of Reasons

Itai Sher*

October 1, 2016

Abstract

Whereas economic models often take preferences as primitive, at a more fundamental level, preferences may be grounded in reasons. This paper presents a model of reasons for action and their weight. The weight of a reason is analyzed in terms of the way that the reason alters the comparison between a default and an alternative. I analyze independent reasons and characterize conditions under the weights of different reasons can be added. The introduction discusses the relevance of reasons to welfare economics.

1 Introduction

Preferences are fundamental to economics in general and normative economics in particular. Economists typically think about decision-making in terms of the maximization of preferences. Alternatively, one could think about decision-making in terms of reasons. Many philosophers have written about the role of reasons in decision-making. Rather than starting with fixed preferences, one can think of people as weighing reasons for and against different actions, and deciding, on the basis of the weight of reasons which action to take.

This paper provides a formal model of decision-making on the basis of reasons.[1] The model is normative rather than descriptive. I do not aim to capture the way that most people reason most of the time. Rather, I aim to provide a way of thinking about how reasons support actions. To get a clearer picture of the contrast between normative and descriptive questions in this domain, contrast the question, "Is R a compelling reason to

*Department of Economics, University of California, San Diego, email: itaisher@gmail.com. I am grateful to Richard Bradley, Chris Chambers, Franz Dietrich, Christian List, Anna Mahtani, Michael Mandler, Marcus Pivato, Peter Sher, Shlomi Sher, Eran Shmaya, Joel Sobel and participants at the Workshop on Reasons and Mental States in Decision Theory at LSE.

[1]For alternative formal approaches, see Dietrich and List (2013) and Dietrich and List (2016).

do a?" with the question "Are people typically compelled by R to do a?". The former is normative, the latter descriptive. Normative models of decision-making are not at all foreign to economics. Expected utility theory, for example, can be taken as a descriptive theory that makes claims about how people actually behave, or as a normative theory that makes claims about how people should behave under uncertainty.

As I am here presenting this paper at a conference on welfare economics, I will begin by motivating the project in terms of what I believe to be its significance for welfare economics. In Section 1.1, I will argue that a solid foundation for welfare economics must address not only the preferences that people have but also the reasons that people take to justify those preferences, at least on those occasions when people's preferences are supported by reasons.[2] With this motivation in hand, in Section 1.2, I will provide a more narrowly focused introduction to the model of reasons that I develop here. I believe that an explicit application of either this model or at least its subject matter to the concerns of welfare economics is an important topic for future research.

1.1 Reasons and Welfare Economics

This section discusses the relevance of reasons to welfare economics. Why should it matter whether we adopt the conceptual scheme and terminology of reasons rather than – or in addition to – that of preferences when we are discussing welfare economics? I will focus on the aggregation of individual values into social values.

A central goal of economics is to find the best ways of satisfying multiple objectives.[3] Each person has his or her own objectives and there are scarce means for achieving those objectives. Because of scarcity, not every objective can be fully satisfied. Therefore, we seek a compromise. Ideally the compromise will not involve any waste. This means that if there is some way of better achieving some objectives without compromising others, this alternative should be seized. In essence, this is the Pareto criterion. Typically, we cannot restrict attention to the Pareto criterion. First, very bad outcomes can be Pareto efficient. For example, suppose that one person's objective is maximized at the expense of everyone else's; all resources are devoted to the realization of this one person's objectives before any resources are devoted to anyone else's. Second, the Pareto criterion is not robust. It only takes one person who prefers x to y whenever everyone else prefers y to x for the Pareto criterion to become completely vacuous. We can call this *the problem of one contrary person*.

[2]This involves a presumption of rationality, at least to some degree. Alternatively, we can view welfare economics as applying – in this case and in many others but not always – to people who are at least somewhat ideally rational.

[3]My formulation of evaluation in welfare economics in this section has been influenced by Dasgupta (2001), specifically by Chapter 1.

Third, and most importantly, for the typical choices that are presented to society, x and y, it is not the case that everyone prefers x to y or everyone prefers y to x. So the Pareto criterion is silent on the most important comparisons.[4] However, that does not mean that all such choices are equally difficult to decide; we may have a sense that some such decisions are much better than others, and we should attempt to elaborate on why we think this and what criteria we are appealing to, which necessitates going beyond the Pareto criterion.

The above considerations imply that if we are to evaluate institutions and policies, we must do so on a basis stronger than that of unanimous agreement. A general way of represent the problem is by means of a *social welfare function*. Because the term "welfare" itself carries strong connotations, and I wish to make only weak assumptions at the moment, I will speak instead of a *social value function*. Let X be a set of outcomes. Then a social value function $v : X \to \mathbb{R}$ ranks those outcomes. That $v(x) > v(y)$ means that x is better than y according to v.

The goal mentioned above was to find a compromise among multiple conflicting objectives supplied by individuals. So far, we have assumed nothing about v that forces it to play that role; v may "impose" values from the outside. To link social and individual values requires that we introduce individual values: Let $v_i : X \to \mathbb{R}$ be individual i's value function, where there are n individuals i, numbered from 1 to n. I use the term "value function" rather than "utility function" because it is less loaded. One way to interpret the mandate that the social value function be based on individual objectives it to posit that the social value function is *individualistic*, meaning that there exists a function f such that

$$v(x) = f(v_1(x), \ldots, v_n(x)). \tag{1}$$

(1) says that social value is a function of individual values. Given that the evaluator is not satisfied to restrict attention to Pareto comparisons, and aggregates the values of different agents according to some function f, by the choice of this function f, she must herself impose certain value judgments.[5] To posit that v is individualistic in the sense of satisfying (1) limits the imposition in that it implies that once people have supplied their inputs, the evaluator must make comparisons according to a *rule* f that relies solely on these inputs, and has no further freedom to make judgements going outside the rule. The evaluator cannot say something like, "apart from what the people have said, I view this outcome as bad, so I will adjust the evaluation downwards." This formalizes a particular sense in which the social judgment is *based* on individual judgments.

[4] Economists sometimes appeal to compensation criteria in such circumstances. As is well known, these criteria are highly problematic. I do not have room to go into these problems here.

[5] At this point, it is natural to ask whether the v_i's are ordinal or cardinal, and also what is being assumed regarding interpersonal comparisons. I will return to this shortly.

To spell out the nature of the social evaluation being discussed here, it is important to specify how we interpret the individual value functions v_i. To do so, consider the following potential interpretations of the inequality $v_i(x) > v_i(y)$.

1. i would choose x over y.
2. i prefers x over y.
3. i would have higher welfare under x than under y.
4. i judges that x is better than y.

Item 4 can be further subdivided as follows.

4.a. i judges that x is morally better than y.

4.b. i judges that in light of all considerations, including moral considerations as well as i's personal priorities, x is better than y.

While some people might want to identify some of these different interpretations, it appears that they are all distinct, at least conceptually. In particular, it seems consistent to affirm any one of these while disaffirming any of the others.[6] Some of these interpretations lend themselves more naturally to cardinal interpretations of v_i, and it seems that the aggregation exercise is more promising if the v_i are cardinal. I think that most of these relations can be given a cardinal interpretation, but I will leave this issue aside because I want to focus on the big picture here.

The important point is that the justification for the aggregation (1) will look very different depending on what it is that one thinks is being aggregated. The simplest and most straightforward version of (1) will occur when we take the value functions as representing welfare.

Welfarist Conception. The welfarist conception selects the individual value functions v_i as measuring welfare in an ethically significant sense. This conception justifies (1) via the claim that welfare is what matters, and, moreover, that welfare is *all that matters*.

The welfarist conception appears contrary to the initial liberal motivation for the aggregation exercise presented above. The starting point was that individuals have their own objectives, but due to scarcity, not all of these objectives can be realized simultaneously.

[6] This consideration may not conclusively show that these notions are distinct, but it should at least give us pause before identifying any of them.

So they must be traded off. A stark and minimal way of doing this would be through the Pareto criterion, but that way is too spare and minimal. So we must impose a more precise trade off. Still we would like this trade off to be a way of balancing individual values rather than an imposition of wholly different values.

The problem with the welfarist conception vis-à-vis the liberal motivation is that people themselves may prioritize other goals over the advancement of their own welfare. Parents may prioritize the welfare of their children; environmentalists may prioritizes preservation of the natural environment; libertarians may prioritize freedom; egalitarians may prioritize equality; religious people may prioritize the dictates of God; ambitious people may prioritize their own success, even when it comes at the expense of their welfare; creators may prioritize their creations, which they may view as surviving past their death, and hence transcending their welfare. In sum, there are a wide variety of goals and values, both self-interested and other-regarding, that people may value above their own welfare.

Now it may be that utilitarianism, with utility interpreted as welfare, is a correct moral theory. And in light of this moral theory, aggregating welfare is what one really wants to do. This would be a justification of the welfarist conception. It is important to note, however, that this is a very different justification than the one from which we started, which was that of finding a compromise between people's objectives. It seems that if we want to realize that conception, we had better take the value functions as representing either preferences or some kind of judgment about what is better. The welfarist conception imposes a very definite value, rather than merely aggregating what people take to be important.

Arrow (1951) had in mind something closer to a judgement than to welfare in *Social Choice and Individual Values.* He wrote,

> It is not assumed here that an individual's attitude toward different social states is determined exclusively by the commodity bundles which accrue to his lot under each. It is simply assumed that the individual orders all social states by whatever standards he deems relevant. A member of Veblen's leisure class might order the states solely on the criterion of his relative income standing in each; a believer in the equality of man might order them in accordance with some measure of income equality ... In general, there will, then, be a difference between the ordering of social states according to the direct consumption of the individual and the ordering when the individual adds his general standards of equity. We may refer to the former ordering as reflecting the tastes of the individual and the latter as reflecting his *values*.

I will now argue that it is not as straightforward to aggregate judgments as it is to aggregate welfare.[7] Consider welfare: There is Ann's welfare and there is Bob's welfare.

[7]For related views, see Mongin (1997). For formal work on judgment aggregation, see List and Pettit

From the standpoint of an evaluator, both may be regarded as goods.[8] The evaluator implicitly trades off one against the other, analogous to different consumption goods from the standpoint of the consumer: Given the constraints faced by society, if Ann is to receive an additional unit of welfare, how much will this cost us in terms of Bob's welfare?

But let us contrast this with the aggregation of judgments. Ann believes that the natural environment has an inherent value, and Bob disagrees. Bob believes that the natural environment is only valuable insofar as it contributes to the welfare of people; he definitely denies, in opposition to Ann, that the natural environment has any value on its own. Then is the evaluator to treat the natural environment as an intrinsic, or non-instrumental, good to be traded off against other goods, such as people's welfare? If she does, then she disagrees with Bob. If she does not, then she disagrees with Ann. One possibility would be to take a compromise position. The evaluator might take the natural environment as a good, but assign it less value, relative to welfare – including relative to Ann's welfare – than Ann does. But notice that in so doing the evaluator is performing quite a different exercise than she did in trading off the welfares of the different agents. It is not merely a question to trading off certain uncontroversial goods – people's welfares – but of taking a compromise position on whether something – the natural environment – is an inherent good.

Consider another kind of case. Both Ann and Bob care about future generations, but Ann believes that carbon dioxide emissions contribute to climate change, whereas Bob believes that climate change is a hoax. As a consequence, Ann and Bob favor different policies. The evaluator might take sides between them, or she might choose a policy that is based on an estimate of the probability that carbon dioxide emissions will lead to any given level of future damages that is intermediate between the probabilities assigned by Ann and Bob. Suppose, however, that Ann has studied the issue and Bob has not. Should this affect the evaluator's attitude? Suppose that Ann can present evidence and Bob cannot, or that Ann can present stronger evidence than can Bob. Does this have any bearing? Do the beliefs of the scientific community, or more specifically, the scientists who are actually studying an issue get any additional weight, or does a climate scientist get just one vote, like everyone else?

Suppose moreover that we have a climate scientist and economist. The climate scientist's views about the climate consequences of various courses of action are more accurate, but the economist's views about certain aspects of the costs of different policies are more accurate. Their preferences are determined by their views. Should the evaluator try somehow to aggregate these preferences, or should she try to aggregate the views and select a policy that is best rationalized be the most accurate view?[9]

(2002), Dietrich and List (2007), and List and Polak (2010).

[8]For a conception of welfare economics as a matter of weighing different people's good, see Broome (1991).

[9]The climate scientist and economist seem here to be playing roles as experts rather than citizens or private individuals, whereas we have been talking about aggregating the values of citizens or private individuals.

The point is that when people judge that one alternative is better than another for a reason, or prefer one alternative to another for a reason, the situation for an evaluator who wishes to aggregate the preferences, values, or judgments of different people, is quite complicated. A central issue is that reasons can conflict in different ways. Scanlon (1998) explains this point, writing,

> Desires can conflict in a practical sense by giving incompatible directives for action. I can want to get up and eat and at the same time want to stay in bed, and this kind of conflict is no bar to having both desires at the same time. Judgments about reasons can conflict in this same way. I can take my hunger to be a reason for getting up and at the same time recognize my fatigue as a reason not to get up, and I am not necessarily open to rational criticism for having these conflicting attitudes. But judgments can also conflict in a deeper sense by making incompatible claims about the same subject matter and attitudes that conflict in this way cannot, rationally, be held at the same time. ... I cannot simultaneously judge that certain considerations constitute a good reason to get up at six and that they do not constitute a good reason to do this.

Scanlon is distinguishing here between two kinds of conflict. Indeed, it appears that the term "conflict" is being used here in two different senses. One type of conflict stems from the incompatibility of different actions. Suppose that two reasons bear on my decision and it is not feasible to take both the action most supported by the first reason and the action most supported by the second reason. I view both reasons as compelling but I cannot satisfy both. This first type of conflict is similar to the conflict that occurs in the welfare aggregation context: It is not possible to simultaneously satisfy all the wants of two people. Some of the wants of one person must come at the expense of some of the wants of the other.

The first kind of conflict occurs when different compelling reasons recommend different – and, moreover, incompatible – actions. I call this a *conflict of recommendation*. Under such a conflict, different reasons suggest different standards for measuring actions; we can measure action by a standard implied by one reason, or by another. We may try to resolve the conflict by *integrating* the two standards into a single unified standard. That may be what happens when we weigh reasons.

The second kind of conflict is quite different: In the second kind of conflict, which I call a *conflict of assertion*, the conflicting reasons correspond to incompatible claims. These claims cannot be true simultaneously, or, if, in some cases, one does think they are truth-apt, they

Thus, it may seem that the discussion is confused. However, what the appeal to experts points out is that when we start talking about judgments, the relevant knowledge and justifications are not equally distributed in society, but are rather concentrated in certain parts of society. Secondly, even if we restrict attention to private individuals, knowledge and justification will not be equally distributed among individuals.

cannot be held simultaneously. For example, Ann holds that we should select the policy best for citizens of our country because we should only be concerned with our fellow citizens' wellbeing, whereas Bob holds that we should select the policy best for the world as a whole because we should be concerned with all people's wellbeing. Alternatively, Ann holds that we should restrict carbon dioxide emissions to limit climate change, and Bob holds that we should not do so because climate change is a hoax. Observe that these conflicts can be conflicts of fact (is climate change happening?) or conflict of value (should we be concerned with all people or just fellow citizens?).

I can summarize my position by saying that while aggregation over cases of conflicting recommendations makes sense, it is questionable whether if makes sense to aggregate over conflicting assertions.

Let me retrace the path that we have taken in this introduction. We started by questioning what it is that we should be aggregating in normative economics. I emphasized a liberal motive for aggregation, that is of respecting people's attitudes rather than imposing our own. I pointed out that any aggregation procedure will involve some imposition, but, the liberal motive suggests that we keep this at a minimum. I next pointed out that utilitarian aggregation, or more generally, aggregation of welfare is not in accord with this liberal motive. Under a liberal view, it seems more appropriate to aggregate judgments backed by reasons. Judgments and the reasons that support them can conflict in different ways. The first kind of conflict is a conflict of recommendation. This is analogous to the case were there are different goods recommended on different grounds, and it makes sense to trade off these goods. It makes sense to aggregate judgments that conflict in this way. In contrast to this, there are conflicts of assertion, which involves mutually incompatible claims. Aggregation across such mutually incompatible claims is questionable. Perhaps it can be done, but it has quite a different character than that of weighing and trading off goods.

Perhaps, the notion of aggregating people's objectives was misguided to begin with. We might want to adopt an alternative conception according to which respecting people's preferences and objectives is not to be done by aggregating them, but rather by setting up *procedurally* fair institutions that do not unnecessarily interfere with people's choices, that distributes rights and freedoms according to certain principles, that make collective decisions democratically, and that encourage discussion and debate, so that people can persuade one another. The thought that the attitudes that we most want to respect in making social decisions are individual judgments backed by reasons might contribute force to this procedural alternative to aggregation. Whether this is so is beyond the scope of the current paper. However, I hope to have shown that taking the fact that people's judgments and preferences are supported by reasons seriously should have an impact on how we conceive of welfare economics.

1.2 An Introduction to the Model

In this paper, I present a model of reasons and their weight.[10] This second part of the introduction will set stage for the formal model to follow. Section 1.2.1 will explain why we need a model. Section 1.2.2 presents a basic picture of making decisions on the basis of reasons. Section 1.2.3 addresses the general question of what reasons are. Section 1.2.4 will presents the essentials of the paper's approach.

1.2.1 Motivation and Assessment

Before proceeding, I would like to say a world about why we need a model at all and how the model can be assessed. The advantage and disadvantage of a formal model, relative to an informal theory, is that it makes more definite claims about certain details. This is a disadvantage because the constraint that the theory spell out certain details precisely may cause the modeler to make claims that are less plausible than those that we would make according to our best informal understanding of reasons; indeed, some claims are bound to be somewhat artificial. It is an advantage because thinking through certain details can test our understanding and raise new questions that we would not have been lead to consider otherwise. A more specific benefit in this case is that the model links philosophical thinking about reasons and their weight to formal models in economics concerning maximization of expected value. In this way it can help to provide a bridge between the perspectives of philosophers and economists.

As I mentioned in the opening paragraphs, what I present here is a normative model, and one might wonder how such a model is to be assessed. We can assess the model on the basis of whether it matches our intuitions and understanding about this subject matter. The model makes some predictions about the relations of the weights of related reasons for different decisions. These are not predictions about behavior, but rather predictions about our normative judgments. We can assess these predictions to see whether they are consistent with our intuitive understanding. As a normative theory, expected utility theory is also assessed on the basis of its fit with our normative judgements.

1.2.2 Weighing Reasons: A Picture

It is useful to present an intuitive picture of what it is to weigh reasons in a decision. 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 the table below.

[10] For a recent collection on essays on the weighing of reasons, see Lord and Maguire (2016).

Reasons For	Weight
R_1	w_1
R_2	w_2
R_3	w_3

Reasons Against	Weight
R_4	w_4
R_5	w_5
R_6	w_6

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:

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

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. One question to be addressed below is when such adding of weights is legitimate.

1.2.3 What are Reasons?

Imagine 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*. These are some examples, but on a deeper level, we can inquire about the nature of (normative) reasons. Opinion is split on this. Raz (1999) writes, "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."[11] 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." Scanlon (1998) writes, "Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favor of it."

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

There are clearly similarities as well as differences among the above accounts. Normative reasons may be conceived of as explanations, as evidence, or as justifications of the proposition that one ought to perform a certain action. Because the goal of this paper is to present a formal framework for modeling 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. The justificatory interpretation is most in line with the way I speak of reasons below. The evidential interpretation fits well with the formalism that I present here. However, because this interpretation is philosophically contentious, and I have some reservations about it myself, I discuss a formal device that could be used to resist this interpretation in Section 2.2.2 (see, in particular the discussion of the set $\mathscr{G}$ of potential reasons).

Formally, I think of reasons as facts of true propositions. I refer to propositions, without specifying whether they are true or false, as potential reasons. Sometimes I speak loosely, using the term "reasons" rather when "potential reasons" would be better. This should not cause any confusion. In this paper, I allow that reasons to be descriptive propositions (e.g., It would make Jim happy if I went to the movies, I told Jim that I would go to the movies with him, etc.) or normative propositions (e.g., it is important to keep one's promises). The model applies more straightforwardly to reasons that are descriptive propositions. Despite the fact that the model assigns probabilities, typically strictly between zero and one, to potential reasons, and it may not seem that this makes sense for normative propositions, I believe that it is legitimate to apply the model to normative propositions as well. Section 4.3 discusses this in the context of a specific example.

1.2.4 This Paper's Approach

I now explain the approach that I take in this paper to modeling reasons and their weight. The basic idea is that the weight of a reason measures the effect that taking the reason into account has on a decision. The model is *doubly comparative*: We compare the decision between an action and a default in the situation when some reason R is taken into account to the decision when R is not taken into account. The first comparison is between the action and the default. The second comparison concerns how the first comparison changes when R is taken into account.[12]

The model relates reasons to value. Either weight or value can be taken to be primitive and the other derived. If value is taken as primitive we think of reasons as affecting the values of actions and thereby derive their weight. Alternatively, we can take the weight of reasons as primitive, and derive implicit values from the weight of reasons. This is elaborated in Section 5.

[12]For comparativism in rational choice, see Chang (2016).

Probability also plays an important role in the theory. We want to compare a decision when a reason R is taken into account to the decision when the reason is not taken into account. What does it mean to not to take R into account? As I explain in Section 4, it often does not make sense to think of not taking R into account as being unaware of the possibility of R. Formally, I model not taking R into account as being uncertain whether R is true, and taking R into account as knowing that R is true. These sorts of comparisons are not important just when assessing the weight of one reason but also in looking at the weights of multiple reasons and asking how they interact. For, example, when can we add these weights? Just as, in the theory of decisions under uncertainty, probabilities allow us to aggregate the different effects of our actions in different states of the world, in the theory presented here, in order to assess the weight of a single reason, we must aggregate over circumstances in which various other combinations or propositions – themselves potential reasons – are true and false. Probability plays this role.

That being said, I give probability a non-standard interpretation, which I call *deliberative probability*. This does not measure your literal uncertainty. Imagine that you know that R is true but to consider the weight of R, you consider how you would decide if you did not take R into account. I treat this as a situation in which you suspend your belief in R and act as though you are uncertain whether R is true. Is not taking R into account more similar to making the decision knowing that R is true or knowing that R is false? This is what deliberative probability measures. Section 4 elaborates on deliberative probability.

Practical decision-making can be thought of normatively in terms of weighing reasons. Another perspective comes from decision theory. A typical formulation is that we should take the action that maximizes value, or expected value, given one's information. Value may be interpreted broadly to include – if possible – all relevant considerations, including moral considerations.[13] The paper presents a model in which weighing reasons is *structurally* the same as maximizing value or expected value given one's information. As I have mentioned, the interpretation of probability, and hence of updating on information and of expected value are different. However, it is useful to have a translation manual between the languages of these two approaches. The relation between weighing and maximizing is spelled out in Section 3.3, and in particular, Proposition 1.

One focus of this paper is the *independence of reasons*. Weighing reasons takes a particularly simple form when it is possible to add the weights of different reasons, as suggested by the example of Section 1.2.2 above. This is not always possible as reasons are not always independent. Section 3 defines the independence of reasons formally and studies this notion.

[13]The qualification "if possible" stems from the possibility that certain moral obligations can not be encoded in terms of the maximization of value; such obligations may differ structurally from any sort of maximization.

1.2.5 Outline

An outline of the paper is as follows. Section 2 presents the model of reasons and their weight. Section 3 studies the independence of reasons, discussing when weights of different reasons can be added. Section 4 discusses deliberative probability. Section 5 shows that weight and value are interdefinable: weight can be defined in terms of value or value in terms of weight. Section 6 concludes.

2 Modeling Reasons and their Weight

This section presents a model of reasons and their weight. Section 2.1 presents what I call the "direct model". This model is minimal. It posits that reasons have weights and discusses the interpretation of those weights. Section 2.2 presents the comparative expected value (CEV) model. This provides one possible foundation for the direct model. It is a more specific model in the sense that it implies that the weights of reasons should have various properties that are not implied by the direct model.

2.1 The Direct Model

In what follows, I will model potential reasons as propositions (or events). Let Ω be a set of states or possible worlds. A proposition R is a subset of Ω.

Suppose that I am contemplating the decision between a default decision d and an alternative a. Let $R_1, \ldots, R_n$ is a set of reasons potentially relevant to my decision.

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 did 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.

However, many decisions – most decisions – involve a mix of knowledge and uncertainty. In such cases, we reason not only about the reasons we have, but also about potential reasons that may emerge. We may ask not only, "What is the bearing of this fact?" but also "How would coming to learn that such and such is true affect my decision?" This means that actual uncertainty is also important, and we must actually integrate deliberative and actual uncertainty. This will be discussed in Section 4.4. Now, I set aside actual uncertainty and focus on deliberative uncertainty.

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

If I were to learn only that proposition R_j held, this might break the indifference. Perhaps R_j will shift the decision in favor of a. Let us imagine that we could quantify *how much better* I should take a to be than d upon learning R_j. We are to imagine here that I learn *only* that R_j obtains and nothing else. Let $w_a(R_j)$ be a real number that represents how much better I should take it to be to select a rather than d *overall* where I to learn R_j (and *only* R_j).

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

If R_j is a reason for a over d, then R_j will shift the decision toward a, and this is represented by the condition that $w_a(R_j) > 0$. If R_j is a reason against a, then R_j will shift the decision away from a, and $w_a(R_j) < 0$. The absolute value $|w_a(R_j)|$ represents the *magnitude* of the reason, that is, how much better or worse a should be taken to be than d; the sign of $w_a(R_j)$ – that is whether $w_a(R_j)$ 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_j a number $w_a(R_j)$. Then we may *define* R_j to be a **reason for** a over d if $w_a(R_j) > 0$, and R_j to be **reason against** a over d if $w_a(R_j) < 0$. We may define $w_a(R_j)$ to be the **weight** of the reason R_j for or against a.

Observe that the the notion of weight that we have adopted here is essentially *comparative*.

It reflects how R_j shifts the comparison between a and d, not how R_j changes the evaluation of a in an absolute sense. The reason for taking a comparative approach will be explained in Section 2.2.5 below in the context of the more detailed CEV model.

Above, I assumed that I learned that only one of the propositions R_j held, but a large part of the importance of thinking about the weight of reasons comes from situations where there are multiple reasons bearing on my decision and I must weigh them against one another. Suppose, for example, that I were to learn that both R_j and R_k held, and that that was all that I learned. Then, this would also shift the comparison of a and d in some way. $w_a(R_j \cap R_k)$ represents the way in which the comparison between a and d would be shifted by this knowledge. Here $R_j \cap R_k$ is the conjunction of R_j and R_k (or more formally, the intersection of R_j and R_k).

Imagine that $w_a(R_j) > 0$. If I had only learned that R_j held, then $w_a(R_j)$ would have represented how much better it would have been overall, from that epistemic standpoint, to have taken action a rather than d. In this case, R_j would have been *decisive*. However, it may be that $w_a(R_j \cap R_k) < 0$, so that having learned R_j *and* R_k, it would have been better to stick the with default. So, assuming that learning R_k does not change the valence of R_j, once I learn both R_j and R_k, R_j still argues for a, but R_j is no longer decisive. Accordingly, the model presented here is a model of *pro tanto* reasons: These are reasons that speak in favor of certain alternatives, but that can be defeated by other reasons. By combining all pro tanto reasons, one comes to an overall judgment about what one should do; so together, all reasons are decisive. In this sense, the model bears on *pro toto* reasons, or, in other words, *all things considered* reasons, as well.

I conclude this section by mentioning two ways in which the assignment of weight can be broadened, both of which will be important for what follows. First, conjunction is not the only relevant logical operation. Let $\bar{R}_j$ be the negation of R_j (formally, the complement of R_j). Then $w_a\left(\bar{R}_j\right)$ represents how the comparison between a and d should change were one to learn that R_j does not hold. Intuitively, $w_a(R_j)$ and $w_a\left(\bar{R}_j\right)$ are of opposite sign. So if $w_a(R_j) > 0$, it is intuitive to expect that $w_a\left(\bar{R}_j\right) < 0$: If R_j is a reason *for* choosing a over d, then $\bar{R}_j$ is a reason *against* choosing a over d. We will return to this.

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

2.2 The Comparative Expected Value Model

2.2.1 The Need for a More Definite Model

The purpose of the "direct model" of the previous section was to provide a more precise picture of the weight of reasons than ordinarily emerges in informal discussions of the topic. The direct model still leaves many details imprecise. For example, how precisely are we to think of the state of knowledge in the position of deliberative uncertainty? The agent does not know whether R_j obtains, but does she have any attitude toward R_j? Is she simply unaware of the possibility of R_j? If she is aware of the possibility, does she think R_j is likely? Does her attitude toward R_j change when she learns that R_k obtains? What properties should the weight function w_a have? I said that it is intuitive that $w_a(R_j)$ and $w_a\left(\bar{R}_j\right)$ should have opposite signs, but can this be provided with a firmer basis, and can something more precise be said about the relationship? What can we say about the relationship between $w_a(R_j), w_b(R_k)$, and $w_a(R_j \cap R_k)$? Between $w_a(R_j), w_b(R_j)$, and $w_{a,b}(R_j)$? When we take R_j into account, do we update the weight of R_k? If so, how? Why should we think of weight in a comparative rather than an absolute sense?

In this section, I present a model of reasons in terms of more familiar conceptual tools. Using these tools, many of the above questions will have clear and well-defined answers. The logic of reasons and their weight will thereby be clarified. The model is useful because it provides a well-defined structure. When specific claims about the structure of reasons and their weight are made, it may be instructive to check whether these claims are either implied by or compatible with the structure posited by the model. In this sense, the model can serve as a benchmark. A second advantage of the model is that it posits a precise relation between reasons and value, and thus suggests an integrated understanding of these two important normative concepts.

A potential disadvantage of the model is that it makes strong assumptions, not all of which have independent intuitive support. But this feature also has merit. A model forces one to make decisions about the structure of reasons, most importantly, with regard to questions that do not have obvious answers. When the answers the model provides seem wrong, one can learn a great deal from diagnosing the problem. Without a model, one would not have been forced to confront these issues.

2.2.2 The Ingredients of the Model

Let d be a default action as above. Let A be a set of alternative actions. Let $A_0 := A \cup \{d\}$ be the set of all actions, including the default. Let Ω be the set of states as above.

Value. For each action $a \in A$, let $V_a : \Omega \to \mathbb{R}$ be a function that assigns to each state a real number $V_a(\omega)$. I interpret $V_a(\omega)$ as the overall value of the outcome that would result from taking action a in state ω. I assume that value is measured on an interval scale.

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 that concept, 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.

For any pair of actions a and b, it is useful to define $V_{a,b}(\omega) = V_a(\omega) - V_b(\omega)$ as the difference in value between a and b, and I define $\widehat{V}_a(\omega) = V_{a,d}(\omega)$ as the difference between the value of a and the value of the default. I refer to the collection $V = (V_a : a \in A)$ as the **basic value function**, and to the collection $\widehat{V} = \left(\widehat{V}_a : a \in A\right)$ as the **comparative value function**.

Probability. I assume that there is a probability measure P on Ω. $P(R)$ is the probability of proposition (or event) R. Formally, I assume that there is a collection of events $\mathscr{F}$ such that P assigns each event R in $\mathscr{F}$ probability $P(R)$; I put the customary structure on $\mathscr{F}$ and so assume that that $\mathscr{F}$ is a σ-field. This means that if R belongs to $\mathscr{F}$, then its complement $\bar{R}$ also belongs to $\mathscr{F}$. Likewise, if R_0 and R_1 belong to $\mathscr{F}$, then so does $R_0 \cap R_1$. Indeed, $\mathscr{F}$ is closed under countable intersections. These assumptions imply that $\mathscr{F}$ is closed under countable unions. Let $\mathscr{F}_+ = \{R \in \mathscr{F} : P(R) > 0\}$. That is, $\mathscr{F}_+$ is the set of events that have positive probability.

If we are in the position of actual uncertainty (see Section 2.1), then $P(R)$ can be interpreted to represent the agent's degree of belief in R. In the position of deliberative uncertainty, P is a probability measure that the agent constructs to represent a hypothetical state of uncertainty. I will call such a constructed probability measure a **deliberative probability**. Section 4 below discusses the motivation and interpretation of deliberative probability.

I do not necessarily assume that each proposition F in $\mathscr{F}$ is a potential reason. Perhaps, some proposition R can explain or help to explain why it is the case that I should take action a, whereas another proposition F is mere evidence that I should take action a. If I subscribe to the theory that a reason R must be part of an explanation of what I ought to do, and cannot be mere evidence of what I ought to do, then I might hold that F cannot be a reason to take action a. Let $\mathscr{G}$ be a subset of $\mathscr{F}$.[14] I call $\mathscr{G}$ the set of **potential reasons**. Then it may be that both F and R belong to $\mathscr{F}$, but, of the two, only R belongs to $\mathscr{G}$; only R is

[14] I allow for the possibility that $\mathscr{G} = \mathscr{F}$.

a potential reason. One might go further and assume that the potential reasons depend on the action in question: R might be a potential reason for a but not for b. I might therefore write $\mathscr{G}_a$ for the set of potential reasons for action a. For simplicity, I will not go so far. Also for simplicity, I will assume that $\mathscr{G}$, like $\mathscr{F}$, is a σ-field. I define $\mathscr{G}_+ = \{R \in \mathscr{G} : P(R) > 0\}$ to the the potential reasons that have positive probability.

Independence of Actions and Events. While I have indexed the value function V_a by actions a, I have not indexed the probability measure P by actions. I am assuming that the choice of action a does not influence that probability of events in $\mathscr{F}$. In other words, events in $\mathscr{F}$ are interpreted as being *independent* of actions in A. These events are interpreted as being "prior" to the choice of action. The events in $\mathscr{F}$ might be *temporally* prior, but these events need not literally occur at an earlier time. For example, consider the event R that it will rain tomorrow. If no action a has an impact on the probability of rain, then R may belong to $\mathscr{F}$.

Imagine that action a would harm Bob if Bob is at home and would not harm Bob if Bob is on vacation, whereas action b would never harm Bob. Suppose that the probability that Bob is at home is $\frac{1}{2}$, independently of which action is taken. Then the probability that Bob is harmed is influenced by the choice of action, and so the event *Bob is harmed* cannot belong to $\mathscr{F}$. However the event H that *Bob would be harmed if action a were taken* has probability $\frac{1}{2}$ independently of whether action a or action b is taken; in this case, H happens to coincide (extensionally) with the event that Bob is at home. So the event H, and other similar events, that specify the causal consequences of various actions, can belong to $\mathscr{F}$.

Events such as *Bob is harmed*, whose probability depend on which action is chosen, can be formally incorporated into the model, but I will leave discussion of this issue to a future occasion.

2.2.3 The Model

I now present the comparative expected value (CEV) model of reasons and their weight. As in Section 2.1, the focus is on the situation of deliberative uncertainty, but, in the most basic respects, the actual uncertainty interpretation is similar.

In the position of deliberative uncertainty, I have an estimate of the value of each action. My estimate of the value of a is its expected value $\mathbb{E}[V_a]$. If Ω is finite, $\mathscr{F}$ contains all subsets of Ω, and we identify each singleton set $\{\omega\}$ with its sole member ω, then $\mathbb{E}[V_a] = \sum_{\omega \in \Omega} V_a(\omega) P(\omega)$. In other words, the expected value of action a is a weighted average of the values of a at different states ω, where the weights are the probabilities $P(\omega)$ of those states. I will call this the **discrete case**. More generally, when we do not necessarily assume

that we are in the discrete case, $\mathbb{E}[V_a] = \int_\Omega V_a dP$.[15]

In the direct model of Section 2.1, I assumed that in the position of deliberative uncertainty, all actions were viewed as equally good. In the model of this section, this idea is captured by assuming that[16]

$$\mathbb{E}[V_a] = 0, \quad \forall a \in A_0. \tag{2}$$

This says that the expected value of taking any action a is zero. I will call assumption (2) **initial neutrality**. I will not always assume initial neutrality. If initial neutrality is assumed, then in the absence of additional reasons, all actions are viewed as equally good. If initial neutrality is not assumed, then some actions start off at an advantage relative to others.

Now imagine that for some potential reason R in $\mathscr{G}_+$, I were to learn R and nothing else. Then I would update my evaluation of each action. The value of action a would become $\mathbb{E}[V_a | R]$. In the discrete case, $\mathbb{E}[V_a | R] = \sum_{\omega \in R} V_a(\omega) \frac{P(\omega)}{P(R)}$. More generally $\mathbb{E}[V_a | R] = \frac{\int_R V_a dP}{P(R)}$.

As in Section 2.1, suppose that I am contemplating a choice between the default action d and an alternative action a. In Section 2.1, R was a reason for a over d if it shifted the comparison between a and d toward a. The following definition captures this idea in the CEV model.

Definition 1 *Let R be a proposition in $\mathscr{G}_+$, and let a be an action. R is a **reason for** a if*

$$\mathbb{E}\left[\widehat{V}_a | R\right] > \mathbb{E}\left[\widehat{V}_a\right]. \tag{3}$$

*R is a **reason against** a if*

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

It is useful to rewrite (3) more explicitly. R is a reason for a if

$$\mathbb{E}[V_a - V_d | R] \geq \mathbb{E}[V_a - V_d].$$

In other words, R is reason for a if the comparison between a and the default d becomes more favorable to a once R is taken into account.

[15] I assume that for all $a \in A$, $V_a : \Omega \to \mathbb{R}$ is an integrable function.

[16] There is no substantive difference between (2) and the assumption that $\mathbb{E}[V_a] = r, \forall a \in A_0$, where r is some fixed real number. Indeed, because I am assuming that value is measured in an interval scale, there is no difference at all.

Definition 2 *Let $R \in \mathscr{G}_+$. The* ***weight*** *of R for/against a is*

$$w_a(R) = \mathbb{E}\left[\widehat{V}_a \,|R\right] - \mathbb{E}\left[\widehat{V}_a\right]. \tag{5}$$

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 the two definitions, 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 direct model of Section 2.1.

Again, it is useful to unwind the definition.

$$\begin{aligned} w_a(R) &= \mathbb{E}[V_a - V_d \,|R] - \mathbb{E}[V_a - V_d] \\ &= (\mathbb{E}[V_a \,|R] - \mathbb{E}[V_d \,|R]) - (\mathbb{E}[V_a] - \mathbb{E}[V_d]), \end{aligned} \tag{6}$$

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.

Definitions 1-2 appealed to a comparison with the default. Instead, one could focus on specific comparisons, and say that R is **reason for choosing** a **over** b if $\mathbb{E}[V_{a,b} \,|R] > \mathbb{E}[V_{a,b}]$ and R is **reason against choosing** a **over** b if $\mathbb{E}[V_{a,b} \,|R] < \mathbb{E}[V_{a,b}]$. It follows from the above definitions that if R is a reason for choosing a over b, then R is a reason against choosing b over a (because whenever $\mathbb{E}[V_{a,b} \,|R] > \mathbb{E}[V_{a,b}]$, it must also be the case that $\mathbb{E}[V_{b,a} \,|R] < \mathbb{E}[V_{b,a}]$). We can also define the **weight of** R **for/against** a **over** b as

$$w_{a,b}(R) = \mathbb{E}[V_{a,b} \,|R] - \mathbb{E}[V_{a,b}].$$

In the direct model of Section 2.1, we introduced $w_{a,b}(R)$ and $w_a(R)$ as undefined terms. In the beginning of this section we posed the question of how the weights $w_{a,b}(R)$, $w_a(R)$ and $w_b(R)$ are related. In the CEV model, we can answer this question. Our definitions imply that[17]

$$w_{a,b}(R) = w_a(R) - w_b(R). \tag{7}$$

[17] To see that (7) holds, observe that

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

where the third equality is derived by rearranging terms.

In other words, the weight of R for/against a over b is the difference between the weight of R for/against a and the weight of R for/against b. Observe that (7) must hold independently of the choice of the default. Had we selected a different default, this would have modified the values $w_a(R)$ and $w_b(R)$, but it would not have modified their difference $w_a(R) - w_b(R)$.

Relatedly, the direct model did not establish any strong relation among weights $w_{a,b}(R)$ for a fixed reason R, when the alternatives being compared, a and b, vary. The CEV model implies that

$$w_{a,c}(R) = w_{a,b}(R) + w_{b,c}(R). \tag{8}$$

Is this reasonable? (8) is precisely as reasonable as (7). To see this, observe that we can rewrite (8) as $w_{a,b}(R) = w_{a,c}(R) - w_{b,c}(R)$. So, if we select c as the default, (8) reduces to (7). So if one believes that it is reasonable to hold that the weight of R for/against a over b is the difference between the weight of R for/against a and the weight of R for/against b, and that the weights of R for/against a and for/against b are to be measured relative to a default, then one ought to believe that (8) is reasonable as well.

2.2.4 Examples

This section illustrates the CEV model with some examples. For the purpose of discussing the examples, the following observation will be useful.

Observation 1 *Let both R and $\bar{R}$ be in $\mathscr{G}_+$ and let a be an action. Then R is a reason for a if and only if*

$$\mathbb{E}\left[\widehat{V}_a \,|R\right] > \mathbb{E}\left[\widehat{V}_a \,\middle|\bar{R}\right]. \tag{9}$$

In other words, (9) is an equivalent reformulation of (3). Similalry, R is a reason against a if and only if $\mathbb{E}\left[\widehat{V}_a \,|R\right] < \mathbb{E}\left[\widehat{V}_a \,\middle|\bar{R}\right]$.

The observation provides another equivalent way of defining R's being a reason for a: Namely, we can define R as being a reason for a if and only if the comparative value of a is greater conditional on R than conditional on the negation of R. Observation 1 follows from the law of iterated expectations.

With the observation in hand, let us proceed to an example. Let R be the proposition *it will rain*. Let a be the action *taking my umbrella with me*. Let the default d be *leaving my umbrella at home*. First suppose that it will not rain. If it does not rain, then it would be better not to carry my umbrella, because, other things being equal, it is better not to be burdened with it. In other words,

$$\mathbb{E}\left[V_d \,\middle|\bar{R}\right] > \mathbb{E}\left[V_a \,\middle|\bar{R}\right]. \tag{10}$$

That is, the expected value of leaving the umbrella is greater than the expected value of taking the umbrella if it will not rain. But if it will rain, the inequality is reversed:

$$\mathbb{E}\left[V_a \,|R\right] > \mathbb{E}\left[V_d \,|R\right]. \tag{11}$$

This is because it is worth the burden of carrying the umbrella to avoid getting wet. It follows from (10) and (11) that

$$\mathbb{E}\left[\widehat{V}_a \,|R\right] = \mathbb{E}\left[V_a \,|R\right] - \mathbb{E}\left[V_d \,|R\right] > 0 > \mathbb{E}\left[V_a \,\middle|\bar{R}\right] - \mathbb{E}\left[V_d \,\middle|\bar{R}\right] = \mathbb{E}\left[\widehat{V}_a \,\middle|\bar{R}\right]. \tag{12}$$

Observation 1 now implies that, as one would want, according to Definition 1, the fact that it will rain is a reason to take my umbrella.

In the rain example, rain is sufficient to make the alternative, taking the umbrella, superior to the default of leaving the umbrella at home, and no rain is sufficient to make leaving the umbrella superior to taking the umbrella. Thus, the consideration of whether it rains or not determines whether or not I should take the umbrella. The consideration is in this sense decisive. This decisiveness is not essential to rain's being a reason to take the umbrella. To illustrate this, let us modify the example. Suppose, for example, that carrying the umbrella is so burdensome that it is not worth taking it even if it rains. So (11) no longer holds. However, it may still be the case that it is not as bad to take the umbrella if it rains than if it does not because, at least, if it rains, the umbrella will keep me dry. In this case, (9) would still hold, and hence rain would still be a reason to take the umbrella, albeit not a decisive one.

"Non-consequentialist" reasons Suppose I am considering whether to go to the movies or stay at home. Action a is going to the movies, and staying at home is the default. Let R be the event that I promised you that I would go. Making the promise makes the difference in value between going and not going, $\widehat{V}_a = V_a - V_d$, larger (in expectation). This may be because it raises the value of going – there is a value to honoring my promise – or it may be because it lowers the value of not going – there is a penalty to violating my promise. It doesn't matter much which of these two ways we represent the matter. However, either way, the promise increases the difference $V_a - V_d$, and so Definition 1 implies that the promise is a reason to go to the movies. This shows that the CEV model can handle reasons with a "non-consequentialist" flavor.

2.2.5 Why is the "Being a Reason For" Relation Defined Relative to a Default?

One might wonder why R's being a reason for a is defined in comparative terms, relative to a default. Alternatively, one might consider the alternative definition:

Noncomparative Reasons. R is a reason for a if $\mathbb{E}[V_a|R] > \mathbb{E}[V_a]$.

The noncomparative definition differs from Definition 1 in that the noncomparative V_a has been substituted for the comparative $\widehat{V}_a$. The condition is still comparative in the sense that it compares the value of a when R is assumed to hold to the value of a when R is not assumed to hold. However, the alternative definition is not comparative in the sense that the value of a is no longer compared to the value of the default; it is not *doubly* comparative, as is Definition 1.

Consider the umbrella example. Recall the interpretation of $V_a(\omega)$ as the overall value of the outcome that would result if action a is taken in state ω. Suppose that if it rains, then the value of all options goes down: If I do not take the umbrella, it would be worse for it to rain than for it not to rain, and even if I take the umbrella, it would be worse for it to rain than for it not to rain. Then recalling that a is the action of taking the umbrella, $\mathbb{E}[V_a|R] < \mathbb{E}[V_a]$. So, if R's being a reason for/against a were given the noncomparative definition then rain would be a reason against taking the umbrella. Since we no longer have a default, let us rename the action of leaving the umbrella at home b. It would also hold that $\mathbb{E}[V_b|R] < \mathbb{E}[V_b]$, so that rain would also be a reading against leaving the umbrella. This is highly problematic: Any proposition – such as the proposition that it will rain in the current example – that is bad news regardless of which action I take would be a reason against every action. Similarly, any proposition that would be good news regardless of which action I take, such as, e.g., that it is sunny, would be a reason for every action. A reason for a is not merely something which is good news conditional on taking a; a reason is something which makes the comparison of a to some other action I might take more favorable to a. This is the reason that I favor the comparative definition, Definition 1, to the noncomparative definition presented in this section.

2.2.6 Weight vs. Importance

This section defines a notion of *importance* and contrasts it with *weight*. This notion will be useful below, particularly in Section 4.3.

Definition 3 *Let both R and $\bar{R}$ belong to $\mathscr{G}_+$. The* ***importance*** *of reason R for/against a is*

$$i_a(R) = \mathbb{E}\left[\widehat{V}_a|R\right] - \mathbb{E}\left[\widehat{V}_a\middle|\bar{R}\right]. \tag{13}$$

The importance of a reason for action a measures how much the comparison between a and d changes when we move from assuming that R holds to assuming that R does not hold. In other words, the importance of the reason measures how much difference it makes whether

we assume that R does or does not hold. This interpretation of importance and the relation between weight and importance are illustrated in Figure 1.

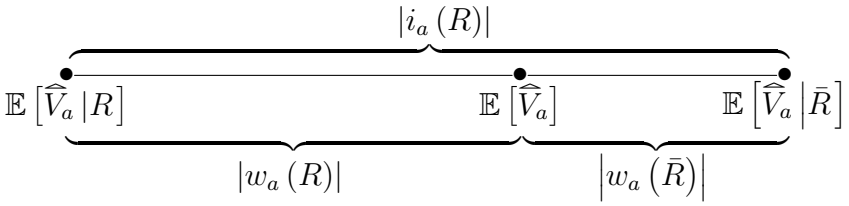

Figure 1: Weight and Importance.

To understand content conveyed by Figure 1, let us consider the relation between weight and importance in more depth. Weight (defined in (5)) and importance (defined in (13)) are quantitative counterparts of the two equivalent qualitative definitions of "being a reason for" (3) and (9), respectively.[18] Definition 3 validates the following relations between weight and importance:

$$i_a(R) = w_a(R) - w_a\left(\bar{R}\right), \tag{14}$$

$$|i_a(R)| = |w_a(R)| + \left|w_a\left(\bar{R}\right)\right|. \tag{15}$$

To understand these equations, suppose that rather than looking at the effect on the decision of learning that R is true relative to a starting position in which we do not know whether or not R is true, we might want to look at the effect of moving from a situation in which we regard R as false to a situation in which we regard R as true; this measures the importance of R. This latter movement can be decomposed into two steps: First we regard R as false, and then we learn that we are mistaken to take this attitude. So we retract the belief that R is false and move to a position of uncertainty. This undoes the effect of learning that $\bar{R}$, and hence moves the comparison by $-w_a\left(\bar{R}\right)$. Then we learn that R is indeed true, and the comparison shifts by $w_a(R)$. The total effect is (14).

With this understanding in hand, let us now turn back to Figure 1. The line segment in the figure represents an interval in the number line. The line segment may be oriented in either direction: Either (i) numbers get larger (more positive) as we move from left to right (in which case $\mathbb{E}\left[\widehat{V}_a \,|R\right] < \mathbb{E}\left[\widehat{V}_a\right]$ so that R is a reason against a) or (ii) numbers get smaller (more negative) as we move left to right (in which case $\mathbb{E}\left[\widehat{V}_a \,|R\right] > \mathbb{E}\left[\widehat{V}_a\right]$ so that R is a

[18] (5) is the difference between the left and right hand sides of inequality in (3), and (13) is the difference between the left and right hand sides of inequality in (9). While (3) and (9) are equivalent in that one is true if and only if the other is true, the quantitative counterparts they generate are different.

reason for a). Updating on R and $\bar{R}$ must move the expectation of V_a in opposite directions from the starting point $\mathbb{E}\left[\widehat{V}_a\right]$.[19] The distance from $\mathbb{E}\left[\widehat{V}_a \middle| R\right]$ to $\mathbb{E}\left[\widehat{V}_a\right]$ is the weight $|w_a(R)|$ (in absolute value), the distance from $\mathbb{E}\left[\widehat{V}_a\right]$ to $\mathbb{E}\left[\widehat{V}_a \middle| \bar{R}\right]$ is the weight $\left|w_a\left(\bar{R}\right)\right|$, and the total distance from $\mathbb{E}\left[\widehat{V}_a \middle| R\right]$ to $\mathbb{E}\left[\widehat{V}_a \middle| \bar{R}\right]$ is the importance $|i_a(R)|$.

While the importance of R and the importance of $\bar{R}$ have opposite signs, reflecting the opposite valence of R and $\bar{R}$, in absolute value the importance of R and $\bar{R}$ the same: Definition 3 implies that

$$i_a\left(\bar{R}\right) = -i_a(R), \tag{16}$$

$$|i_a(R)| = \left|i_a\left(\bar{R}\right)\right|. \tag{17}$$

The importance of R for or against a can be interpreted in terms of the answer to the question

- In choosing between a and d, how important is it whether R or $\bar{R}$ is true?

This question is symmetric between R and $\bar{R}$. To be concrete, suppose that a corresponds to going to the movies with you, and d corresponds to staying at home. Let R be the proposition that I promised you I would go. Then the question concerning the importance of the promise is: "In deciding whether to go to the movies with you, how important is it whether I promised you that I would go?". The answer to this question is the same as the answer to "In deciding whether to go to the movies with you, how important is it whether I did not promise you that I would go?", as expressed by (17).[20]

Now imagine that I did promise you that I would go. Then it is natural to think of my promise as a weighty consideration in favor of going. Imagine instead that I did not promise you that I would go. Then, it would *not* seem natural to treat my *lack* of promise as a weighty consideration against going. Indeed, it seems strained to say that the fact that I did not promise is a reason not to go at all. The fact that I did not promise seems neutral. If one accepts the CEV model, then one has to accept that if R (e.g., promising) is reason for a, then $\bar{R}$ (e.g., not promising) is a reason against a (see Section 5 for a qualification). However, $\bar{R}$ can be represented as a *very minor* reason in the sense that it has a very small weight (in absolute value). This seems to be a reasonable way to represent the situation: Not promising is a reason not to go because, after all, the fact that I did not promise rules out one reason for going without implying any other reason for going or ruling out any reason

[19] It is also possible that R is irrelevant to the decision so that $\mathbb{E}\left[\widehat{V}_a \middle| R\right] = \mathbb{E}[V_a] = \mathbb{E}\left[\widehat{V}_a \middle| \bar{R}\right]$.

[20] The sign is reversal in (16) reflects that we are keeping track of the opposite valence of a reason and its negation.

for not going. Not promising is only a very minor reason because it only rules out one of many possible reasons for going.

The above example illustrates why, in contrast to the natural assumption (17) for importance, it will generally not be the case that $|w_a(R)| = \left|w_a\left(\bar{R}\right)\right|$; the weight of R can be much larger (or smaller) than the weight of $\bar{R}$. This is illustrated in Figure 1, in which the lengths of the two subintervals differ. This is one important aspect of the difference between weight and importance; importance is symmetric between a proposition and its negation, whereas weight is not. The analysis of Section 3 below highlights a further difference: One can add weights, but one cannot add importances. (This, along with the preceding contrast, justifies the choice of terminology.)

3 Independent Reasons

This section focuses concerns the way that different reasons combine. That is to say, how is the weight of a conjunction or reasons $w_a(R_1 \cap R_2)$ related to the weights of the individual reasons $w_a(R_1)$ and $w_a(R_2)$? The simplest, most straightforward relation, would be that the weight of the conjunction is the sum of the weights of the individual reasons. That is, $w_a(R_1 \cap R_2) = w_a(R_1) + w_a(R_2)$. The case in which the weights of the reasons can be added is the case of *independent reasons*. Independent reasons will be the focus of this section. I will first define formally what it means for the weights of different reasons to be independent. I will relate independence of reasons to properties of the associated probability measure and value function in the CEV model.

It is important to emphasize at the outset that independence is not a thesis about reasons. It cannot be the case that reasons are independent of one another *in general*. Rather it may be that *some* specific reasons are independent of some other specific reasons for some actions. For example, R_1 may be independent of R_2 for action a; but R_1 will not be independent $R_1 \cup R_2$; more generally, because reasons stand in logical relations to one another, it will be hopeless to seek for a theory in which reasons are independent in general. To take a concrete example, the facts that (i) an action would cause me to lose at least \$10 and that (ii) it would cause me to lose at least \$100 are not independent reasons not to take the action. In contrast, the fact that an action would help Ann, and the fact that it would help Bob may be independent reasons. Note finally, that independence is relative to the action under consideration. R_1 and R_2 may be independent reasons for action a, but at the same time, R_1 and R_2 may not be independent reasons for b. This will be illustrated below.

Sometimes, when justifying an action, there are multiple logically equivalent arguments for that action, some of which appeal to independent reasons and others of which do not. This is because it is possible to carve up the reasons for an action in different ways, separating them

into either independent parts or non-independent parts. An interesting question that arises is: Are justifications for actions in terms of independent reasons superior to justifications in terms of non-independent reasons? This is discussed in Section 3.5.

Before proceeding, some preliminaries are required.

3.1 Preliminaries

Throughout this section (I mean all of Section 3), I fix a weight function $w = (w_a : a \in A)$, where for each $a \in A$, $w_a : \mathscr{G}_+ \to \mathbb{R}$, and an associated pair (V, P) of a value function and probability measure. Definition 2 of Section 2.2.3 showed how a weight function w can be derived from a pair (V, P) via (3). Section 5 below shows how a pair (V, P) can be derived from a weight function w. So we can think of either w or (V, P) as primitive and the other as derived.

3.1.1 Weights of Collections

Say that a collection $\mathscr{R} = \{R_1, \ldots, R_n\} \subseteq \mathscr{F}$ of reasons is **consistent** if $P\left[\bigcap_{i=1}^n R_i\right] > 0$. An event F is **consistent with** $\mathscr{R}$ if $\mathscr{R} \cup \{F\}$ is consistent. Say that $\mathscr{R}$ is **logically independent (LI)** if $\forall I \subseteq \{1, \ldots, n\}$, $\{R_i :\in I\} \cup \{\bar{R}_i : i \in \bar{I}\}$ is consistent, where $\bar{I} := \{1, \ldots, n\} \setminus I$.[21]

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:

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

In other words, the weight of a collection of reasons is just the weight of their conjunction (or, more formally, their intersection). The hat $\hat{}$ differentiates weight applied to a collection of reasons from weight applied to a single reason.

3.1.2 Conditional Weights

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

[21] Of course, in its details, this definition 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. This is a weak form of independence, weaker, than, say, causal independence or probabilistic independence.

(see Sections 2-4), but from an intermediate deliberative position in which some of these reasons have 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:

$$w_a\left(R_0 \,|\mathscr{R}\right) = \mathbb{E}\left[\widehat{V}_a \,\middle|\, R_0 \cap \left(\bigcap_{i=1}^{n} R_i\right)\right] - \mathbb{E}\left[\widehat{V}_a \,\middle|\, \bigcap_{i=1}^{n} R_i\right]. \tag{18}$$

That is to say, $w_a\left(R_0 \,|\mathscr{R}\right)$ is the difference in the expected value of a that conditioning on R_0 makes when one has already conditioned on the reasons in $\mathscr{R}$. For any proposition $R_1 \in \mathscr{G}_+$, I write $w_a\left(R_0 \,|R_1\right)$ instead of $w_a\left(R_0 \,|\{R_1\}\right)$, so that, in particular, $w_a\left(R_0 \,|\mathscr{R}\right) = w_a\left(R_0 \,|\bigcap_{i=1}^{n} R_i\right)$. Observe that $w_a\left(R \,|\emptyset\right) = w_a\left(R\right)$. Observe also that

$$w_a\left(R_0 \,|\mathscr{R}\right) = \hat{w}_a\left(\{R_0\} \cup \mathscr{R}\right) - \hat{w}_a\left(\mathscr{R}\right). \tag{19}$$

If one wants to take weights, rather than values and probabilities, as basic, then then one can treat (19) as the definition of conditional weight $w_a\left(R_0 \,|\mathscr{R}\right)$, and treat (18) as a derived property.

3.2 Definition of Independent Reasons

We are now in a position to define the independence of reasons.

Definition 4 *$\mathscr{R}$ is a collection of* ***independent reasons*** *for/against a if and only if for all $R_0 \in \mathscr{R}$ and all for all nonempty $\mathscr{S} \subseteq \mathscr{R}$ with $R_0 \notin \mathscr{S}$,*

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

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

In other words, $\mathscr{R}$ is a collection of independent reasons if conditioning on any subset of reasons does not alter the weight of any reason outside the set.

I will find it convenient to work with a stronger notion of independence. Let $\mathscr{R} = \{R_1, \ldots, R_n\}$ be a collection of *logically* independent reasons (see Section 3.1.1). If $\mathscr{T} \subseteq \mathscr{R}$, let

$$\mathscr{R}_{\mathscr{T}} := \{R : R \in \mathscr{T}\} \cup \{\bar{R} : R \in \mathscr{R} \setminus \mathscr{T}\}.$$

That is, $\mathscr{R}_{\mathscr{T}}$ is the set of reasons that result by "affirming" the reasons in $\mathscr{T}$ (treating these as true) and "disaffirming" the rest (treating these as false). For example, if $\mathscr{R} = \{R_1, \ldots, R_5\}$ and $\mathscr{T} = \{R_1, R_3, R_4\}$, then $\mathscr{R}_{\mathscr{T}} = \{R_1, \bar{R}_2, R_3, R_4, \bar{R}_5\}$.

Definition 5 *$\mathscr{R}$ is a collection of **strongly independent reasons** for/against a if and only if for all $\mathscr{T} \subseteq \mathscr{R}$, $\mathscr{R}_{\mathscr{T}}$ is a-independent. To abbreviate, we say that $\mathscr{R}$ is **strongly a-independent**.*

Thus, a collection of reasons is strongly a-independent if, negating any subset to the reasons, the collection remains independent. 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 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, or even a signed measure – it is not additive[22] – and strong a-independence is indeed a property that is harder to satisfy than a-independence. Appendix B presents an example illustrating this.

3.3 Weighing Reasons and Maximizing Expected Value

This section establishes a relationship between weighing reasons and maximizing expected value. In the CEV model, these two apparently quite different visions of decision-making have the same structure. In the case in which reasons are independent, adding the weights of reasons coincides with maximizing expected value.

Let us return to the position of deliberative uncertainty. Imagine that, initially, at time 0, you know nothing. Then at time 1, you will learn various facts. In particular, recall that $\mathscr{G}$ is the set of potential reasons. Let $\mathscr{I} \subseteq \mathscr{G}$. I assume that $\mathscr{I}$, like $\mathscr{G}$, is a σ-field (i.e., closed under complementation and countable intersection). The propositions in $\mathscr{I}$ are precisely those that you will learn, if true, at date 1. That is, for all propositions R in $\mathscr{I}$ that are true at state ω, (i.e., $\omega \in R$) you will learn R at date 1 in ω if and only if $R \in \mathscr{I}$. You will never "learn" any false propositions. What you learn depends on the state ω. So $\mathscr{I}$ represents your information. For simplicity, I assume that $\mathscr{I}$ contains only countably many propositions and each nonempty $I \in \mathscr{I}$ has nonzero probability: $P(I) > 0$. Let $\mathscr{I}(\omega)$ be the conjunction of all propositions that you learn at ω. Formally, $\mathscr{I}(\omega) = \bigcap \{I \in \mathscr{I} : \omega \in I\}$. The above assumptions imply that $P(\mathscr{I}(\omega)) > 0$ for all $\omega \in \Omega$. Observe that $\mathscr{I}(\omega)$ is itself a proposition in $\mathscr{I}$.

We imagine that the state is chosen according to the probability measure P. After receiving your information at date 1, you choose an action from the set A_0, which includes the default action d, and all alternative actions. In fact, it is not important that the state is actually chosen according to P. What matters is only that at state ω, you evaluate the expected value of actions in the same way that you would if the state had been chosen

[22] That is, when R_1 and R_2 are disjoint, it is not generally the case that $w_a(R_1 \cup R_2) = w_a(R_1) + w_a(R_2)$.

according P to and you updated on your information $\mathscr{I}(\omega)$. Thus the expected value that you assign to action a at state ω is $\mathbb{E}[V_a | \mathscr{I}(\omega)]$.

Definition 6 *A consistent collection $\mathscr{R} = \{R_1, \ldots, R_n\} \subseteq \mathscr{I}$ of reasons is a-**complete** if for all $I \in \mathscr{I}$ that are consistent with $\mathscr{R}$ (i.e., $P[I \cap (\bigcap_{i=1}^{n} R_i)] > 0$), $w_a(I|\mathscr{R}) = 0$.*

In other words, a collection of reasons is a-complete if there are no other potential reasons (in $\mathscr{I}$) that have any weight for a conditional on the reasons in the $\mathscr{R}$.

For each action $a \in A$, define $\delta_a = \mathbb{E}[V_a]$. We can call δ_a the **default advantage** of action a. δ_a is the value that would be assigned to action a in the absence of any information. If we adopt the position of **initial neutrality** – before acquiring reasons, there is no basis for preferring one action over the other, (see Section 2.2.3), then this amounts to the assumption that $\delta_a = \delta_b$ for all $a, b \in A_0$ (where recall, A_0 also contains the default). For (21) below, observe that I adopt the standard convention that $\bigcap_{j=1}^{0} R_j = \Omega$.

Proposition 1 *Let $\mathscr{R}_a = \left\{R_1^a, \ldots, R_{n_a}^a\right\}$ and $\mathscr{R}_b = \left\{R_1^b, \ldots, R_{n_b}^b\right\}$ be consistent collections of reasons contained in $\mathscr{I}$, and let $R_0 := (\bigcap_{i=1}^{n_a} R_i^a) \cap \left(\bigcap_{i=1}^{n_b} R_i^b\right)$. Assume initial neutrality. Assume further that $\mathscr{R}_a$ is a-complete and $\mathscr{R}_b$ is b-complete. Then,*

$$\forall \omega \in R_0, \qquad \mathbb{E}[V_a | \mathscr{I}(\omega)] \geq \mathbb{E}[V_b | \mathscr{I}(\omega)] \Leftrightarrow \hat{w}_a(\mathscr{R}_a) \geq \hat{w}_b(\mathscr{R}_b). \tag{20}$$

Also,

$$\forall \omega \in R_0, \qquad \mathbb{E}[V_a | \mathscr{I}(\omega)] \geq \mathbb{E}[V_b | \mathscr{I}(\omega)] \Leftrightarrow \sum_{i=1}^{n_a} w_a\left(R_i^a \middle| \bigcap_{j=1}^{i-1} R_j^a\right) \geq \sum_{i=1}^{n_b} w_b\left(R_i^b \middle| \bigcap_{j=1}^{i-1} R_j^b\right). \tag{21}$$

If $\mathscr{R}_a$ is a-independent and $\mathscr{R}_b$ is b-independent, then,

$$\forall \omega \in R_0, \qquad \mathbb{E}[V_a | \mathscr{I}(\omega)] \geq \mathbb{E}[V_b | \mathscr{I}(\omega)] \Leftrightarrow \sum_{i=1}^{n_a} w_a(R_i^a) \geq \sum_{i=1}^{n_b} w_b\left(R_i^b\right). \tag{22}$$

What the result shows is that when you choose by weighing reasons, then, structurally, it is as if you chose by maximizing expected value conditional on the information that you received. We imagine that the above described updating occurs in the position of deliberative uncertainty. One could also give this scenario the interpretation of actual uncertainty. Then the interpretation would be that if you maximize expected value conditional on your

information, then structurally, it is as if you are weighing reasons.[23] I view the value of this result as resting in the translation manual it provides between the languages of maximizing expected value and of weighing reasons.

The result suggests the following three step decision procedure: (i) for each action a, identify a set $\mathscr{R}_a$ of reasons that is sufficient to make as complete an evaluation of action a as our knowledge will allow, (ii) determine the aggregate weight of the reasons $\mathscr{R}_a$ for action a relative to a default, using the same default for each action, and (iii) choose the action a with the maximum aggregate weight of reasons.

Suppose that in step (i), we determine that some reason R does not belong to $\mathscr{R}_a$. This does not mean that as a stand-alone reason R has no weight for or against a; it means only that we can find a set of reasons $\mathscr{R}_a$ that is complete for a and that does not include R. Then we do not need to consider R further in evaluating a. It may be that R *does* belong to $\mathscr{R}_b$. So we may use different reasons determine the aggregate weight of reasons in favor of or against actions a and b; having used these different reasons, we need only compare the derived aggregate weights to decide between a and b.

For example, in deciding whether to become a doctor or lawyer, I may appeal to the fact that I am squeamish as a reason against becoming a doctor. My squeamishness may or may not be relevant to my becoming a lawyer, but, in any event, if I find a complete set of reasons for evaluating the prospect of becoming a lawyer that does not include my squeamishness, then my squeamishness will contribute to the aggregate weight in favor or against being a doctor and not to that in favor or against being a lawyer.

Suppose that we accept the decision metric of maximizing expected value subject to available information. Then, Proposition 1 says that adding the weights of the reasons for each of the reasons is a legitimate way of making the better decision *if* the reasons supporting each action are independent for that action (see (22)). If the reasons supporting an action are not independent, it is not legitimate to add their weights.

However, a more general procedure is always valid, whether the reasons supporting actions are valid or not. In this procedure, we add not the unconditional weights but conditional weights. In particular, we perform the following procedure in evaluating action a:

- Start with the unconditional weight $w_a\left(R_1^a\right)$,
- then add the weight $w_a\left(R_2^a \,|R_1^a\right)$ of R_2 conditional on R_1,
- then add the weight of $w_a\left(R_3^a \,|R_1^a \cap R_2^a\right)$ of R_3 conditional on R_1 and R_2,

[23] I have suggested above that if we use actual, rather than deliberative, probability, then the weights that we assign may not correspond to the weights that we might intuitively feel should be assigned. The main point here is that the two different views of decision-making are structurally similar.

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

That is, we add the weight of reasons R^a_i one at a time, each time, conditioning on reason just added (in addition to the reasons previously conditioned on) before taking the weight of the next reason R^a_{i+1}. This is the procedure captured by (21) in Proposition 1. Observe that the reasons in $\mathscr{R}_a$ can be added in any order, using this procedure, and we will arrive at the same aggregate weight.

One might think that in assessing the weight of a reason R^a_i for or against action a, we should consider the **marginal contribution** of the reason to overall weight conditional on all other reasons

$$w_a\left(R^a_i \left| \text{all other reasons} \right.\right) = w_a\left(R^a_i \left| \bigcap_{j \neq i} R^a_j \right.\right) \tag{23}$$

That is, rather than retreating to a position in which we suspend knowledge of other operative reasons, to determine its weight, we instead consider the effect of taking reason R^a_i into account holding fixed all other reasons that we know to obtain.

Observe that when reasons are not independent, we cannot add the marginal contributions (23) to arrive at total weight $\hat{w}_a\left(\mathscr{R}\right)$, but when $\mathscr{R}_a$ is a-independent, then

$$w_a\left(R^a_i\right) = w_a\left(R^a_i \left| \bigcap_{j=1}^{i-1} R^a_j \right.\right) = w_a\left(R^a_i \left| \bigcap_{j \neq i} R^a_j \right.\right),$$

so we can arrive at aggregate weight equivalently by adding weights of any of the three forms $w_a\left(R^a_i\right)$, $w_a\left(R^a_i \left| \bigcap_{j=1}^{i-1} R^a_j \right.\right)$, or $w_a\left(R^a_i \left| \bigcap_{j \neq i} R^a_j \right.\right)$.

Proposition 1 assumed initial neutrality. How would the proposition have changed if we had not assumed initial neutrality? (20) would have become

$$\forall \omega \in R_0, \qquad \mathbb{E}\left[V_a \left| \mathscr{I}\left(\omega\right) \right.\right] \geq \mathbb{E}\left[V_b \left| \mathscr{I}\left(\omega\right) \right.\right] \Leftrightarrow \hat{w}_a\left(\mathscr{R}_a\right) - \hat{w}_b\left(\mathscr{R}_b\right) \geq \delta_b - \delta_a.$$

That is, a has a higher expected value than b if and only if the weight of the reasons for a are in excess of the weight of reasons for b by at least the difference between the default advantage of b over a, $\delta_b - \delta_a$. The other parts of the proposition would be modified similarly.

3.4 Chararacterizations

This section presents some characterizations of (strongly) independent reasons in terms of the value function V and probability measure P. Throughout, I hold fixed some set of potential reasons $\mathscr{R} = \{R_1, \ldots, R_n\}$. Let $\mathscr{F}_{\mathscr{R}}$ be the field generated by $\mathscr{R}$. That is, $\mathscr{F}_{\mathscr{R}}$ is the closure of $\mathscr{R}$ under intersection and complementation. Since $\mathscr{F}_{\mathscr{R}}$ contains only finitely many sets, there is no distinction between $\mathscr{F}_{\mathscr{R}}$'s being a field and $\mathscr{F}_{\mathscr{R}}$'s being a σ-field. A random variable is a function $X : \Omega \to \mathbb{R}$ that is measurable with respect to $\mathscr{F}$. For any random variable X, define $X^{\mathscr{R}} = \mathbb{E}[X \,|\mathscr{F}_{\mathscr{R}}]$. For a formal treatment of measurability and for what it means to condition a random variable on a σ-field, see Sections 13 and 34 of Billingsley (1995). For any event $R \in \mathscr{F}$, 1_R is the indicator function that has value 1 in states $\omega \in R$ and value 0 in states $\omega \notin R$.

I will now state three joint conditions on V, P, and $\mathscr{R}$, which I will call **independence conditions.** Within each of these conditions, when I use the term "independence", I mean probabilistically independence with reference to the probability measure P.

IC1 There exists random variables U_i, $i = 1, \ldots, n$ such that:

1. $\widehat{V}_a = \sum_i U_i$, and
2. $U_i^{\mathscr{R}}, R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_n$ are mutually independent for $i = 1, \ldots, n$.
3. $U_i^{\mathscr{R}}, R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_n$ are mutually independent conditional on R_i and conditional on $\bar{R}_i$ for $i = 1, \ldots, n$.

IC2

1. $\widehat{V}_a^{\mathscr{R}} = \sum_{i=1}^n w(R_i)\, 1_{R_i} + \sum_{i=1}^n w\left(\bar{R}_i\right) 1_{\bar{R}_i} + \mathbb{E}[V_a]$
2. $R_1, \ldots, R_n$ are mutually independent.

IC3 There exists random variables Z_i^j, $i = 1, \ldots, n, j = 0, 1$ such that:

1. $\widehat{V}_a = \sum_{i=1}^n Z_i^1 1_{R_i} + \sum_{i=1}^n Z_i^0 1_{\bar{R}_i}$
2. $\left(Z_i^j\right)^{\mathscr{R}}, R_1, \ldots, R_n$ are mutually independent for $i = 1, \ldots, n$ and $j = 0, 1$.

The following definition is useful: $\mathscr{R}$ **contains an a-empty reason** if there exists i such that for all $J \subseteq \{1, \ldots, n\} \setminus i$, $w_a\left(R_i \,|\bigcap_{j \in J} R_j\right) = 0$. An a-empty reason is a reason that has no weight, no matter which other reasons we condition on.

Proposition 2 *Suppose that $\mathscr{R}$ contains no a-empty reasons. Then the following conditions are equivalent:*

1. *$\mathscr{R}$ is strongly a-independent.*
2. *V, P, and $\mathscr{R}$ jointly satisfy IC1.*
3. *V, P, and $\mathscr{R}$ jointly satisfy IC2.*
4. *V, P, and $\mathscr{R}$ jointly satisfy IC3.*

Observe that condition 1, like the other conditions, depends on V and P as well as $\mathscr{R}$. Each of the independence conditions IC1-IC3 have a similar flavor: (i) The value function can be represented as a sum of functions, each of which depends on only one of the reasons R_i, and (ii) the different reasons are probabilistically independent. Intuitively, (ii) means that updating on one of the reasons does not convey information about any of the other reasons.

3.5 An Example

Suppose that I am a utilitarian. Suppose moreover that I am considering a project that would be good for Ann but bad for Bob. My value function might then be the sum of Ann's utility U_A and Bob's utility U_B. Using Proposition 2 and specifically the equivalence of independence of reasons and IC1 to justify treating the effect on Ann and the effect on Bob as independent reasons. This example, and in particular the technical details, are discussed in Appendix C.

Here, I just note a few interesting aspects of the example. Suppose that the effects of the project can be decomposed into effects on the agent's wealth and on their health. Then instead of arguing for or against the project by appeal to reasons concerning the agent's utilities, we can argue by appeal to reasons concerning the aggregate effects on wealth and the aggregate effects on wealth. Under certain assumptions, the aggregate effects on wealth will constitute reasons that are independent of the aggregate effects on health.

This highlights that we may sometimes be able to argue for courses of action in different ways. We can appeal to reasons about utilities or reasons about wealth and health. Each sets of reason may cut up the world in different ways. At the same time, each set of reasons may be redundant in the presence of the other, and yet each individually may be part of an argument for or against the course of action.

It may sometimes happen that two sets of reasons are equivalent in terms of the information they convey but cut up the world in a different ways. One argument is in terms of independent reasons and the other is not. This raises the question of whether the argument

in terms of independent reasons is in some sense better and to be preferred. This is discussed further in Appendix C.

3.6 Quasi-weights and Neutral Defaults

I have emphasized above that a collection of reasons might be independent for one action a while not being independent for another action b. Recall however that weights are defined in terms of a second order difference, involving both a variation in action (from a to d) and a variation in whether the reason R is known to hold. This is expressed explicitly in (6). So for $\mathscr{R}$ to be a-independent involves a relation of $\mathscr{R}$ to both V_a and V_d. I now briefly consider how independence might be defined in relation to V_a only.

Inert default. Say that the default is **inert** if $V_d(\omega) = 0$ for all $\omega \in \Omega$.[24] In other words, the default is inert if it is an action that has the same value in every state of the world. In this case

$$w_a(R) = \mathbb{E}[V_a | R] - \mathbb{E}[V_a], \tag{24}$$

so that the weight of a reason depends only on V_a, and indeed we can replace all instances of $\widehat{V}_a$ by V_a above. This is the simplest case in which a-independence can be defined solely in reference to the relation of $\mathscr{R}$ to V_a. Observe that weight is still *defined* comparatively – that is, in comparison to the default – as in (5). (24) is a *consequence* of the inertness of the default, not a definition. Accordingly, the objection to noncomparative definition of weight raised in Section 2.2.5 does not apply.[25]

Quasi-weights. Define the **quasi-weight** of a reason to be $\tilde{w}_a(R) = \mathbb{E}[V_a | R] - \mathbb{E}[V_a]$. This is not a genuine weight for the reasons set out in Section 2.2.5. For any set $\mathscr{R}$ of reasons and reason R_0, define the conditional quasi-weight of R_0 given $\mathscr{R}$ analogously to the way that conditional weights were defined in Section 3.1.2 with V_a playing the role of $\widehat{V}_a$. Say that $\mathscr{R}$ is **quasi-a-independent** if for all $\mathscr{S} \subseteq \mathscr{R}$, and $R_0 \in \mathscr{R} \setminus \mathscr{S}$, $\tilde{w}_a(R_0 | \mathscr{S}) = \tilde{w}_a(R_0)$. This is analogous to Definition 4. One can prove a result analogous to Proposition 2 with quasi-weights rather than weights. So while quasi-weights are not genuine weights for the reasons explained in Section 2.2.5, they may still be a good guide to decision-making. In particular, when we have quasi-independent reasons for different actions, we can add the quasi-weights to decide which action is best.

[24] More generally, we can say that the default is inert if $V_d(\omega) = r$ for all $\omega \in \Omega$ for some real number r.

[25] In particular, it is not possible for good news to count as a reason for every action because there is no event that raises the value of the default. It is perfectly sensible that some event which raises the value of every action but the default be a reason to do anything but the default.

4 Deliberative Probability

This section interprets the notion of probability in the CEV model. I elaborate the interpretation that I call *deliberative probability*. This interpretation corresponds to the position of deliberative uncertainty (see Section 2.1).

4.1 Conventional Interpretations of Probability

First, I mention the more conventional interpretation that corresponds to the situation of actual uncertainty. Probability may be interpreted subjectively as giving the agent's degree of belief in various propositions. Probability can also be interpreted as objective probability if states have objective probabilities and the agent knows what these probabilities are. I refer to these conventional interpretations, both the subjective and the objective interpretation, as *prospective* because they are most relevant with regard to future events.[26]

I think that the theory is interesting under this interpretation. If the CEV model is interpreted in this way, then the model is well suited to the evidential interpretation of reasons and their weight. This interpretation is valuable because it provides a translation manual between updating expected value on information, which is well understood, and the weighing of reasons, which is more difficult to grasp. It is valuable to have such a translation manual between these theories even if one thinks that the translation is not perfect.

These advantages notwithstanding, the conventional interpretation is not my preferred interpretation of the model. In this section, I will spell out my preferred interpretation in terms of deliberative probability.

4.2 Interpreting Deliberative Probability, Part 1

Deliberative attitudes. I start with the basic idea of deliberative attitudes. To motivate this idea, consider three reasons that may bear on my going to the movies with Jim:

- R_p : I promised Jim that I would go to the movies with him.
- R_ℓ : Jim loves going to the movies with a friend.
- R_h : Jim is a human (as opposed to a dog).

[26] It is of course possible to discuss objective and subjective probability with respect to past events. For example, if a fair coin was flipped yesterday, then the objective probability of its coming up heads *was* $\frac{1}{2}$ yesterday. We may also recall that our subjective estimate of the chances of rain yesterday was $\frac{1}{3}$.

Suppose I know that R_p, R_ℓ, and R_h are true. I want to assess the weight of R_p for going to the movies with Jim. To assess the specific weight of R_p, I may want to abstract away from my knowledge of other reasons. After all, it is the weight of R_p alone that I would like to determine. But what does this "abstracting away" consist in? Specifically, to assess the weight of R_p, what attitude do I adopt toward R_ℓ and R_h? Consider R_h specifically. In assessing the weight of my promise to go to the movies with Jim, can I really adopt no attitude toward the proposition that Jim is a human? No, in considering R_p, I must form some coherent picture of the situation and so take some attitude toward R_ℓ and R_h. The uncertainty in the position of deliberative uncertainty described in Section 2.1 must incorporate such an attitude.

Deliberative probability I will now develop a picture of what it means to abstract away from other reasons in such a way that a coherent picture of the situation is maintained. Imagine that $R_0, R_1, \ldots, R_n$ is a list of propositions, each of which you *know* to be true, and each of which is a reason for or against action a as opposed to the default d.[27] Imagine further that for the purpose of individually assessing the weight of each of these propositions as reasons, you retreat to a deliberative position in which you suspend your knowledge of these propositions (as in Sections 2.1 and 2.2).

In the CEV model, your attitude toward the truth of $R_0, \ldots, R_n$ in the above deliberative situation is modeled via a *deliberative probability* measure P. The probability measure P represents your *deliberative attitude*. It does not represent your *actual* beliefs. By assumption, you *know* that $R_0, \ldots, R_n$ obtain, but P only assigns these propositions certain probabilities. For the purposes of deliberation, you treat R_0 *as if* you assign R_0 probability $P(R_0)$. $P(R_0)$ is the probability you assign to R_0 when you suspend your belief in R_0, and also suspend your belief in the other reasons $R_1, \ldots, R_n$. You do not "take reasons $R_1, \ldots, R_n$ into account". By not taking these reasons into account, I do not mean that you ignore the possibility that they obtain; I mean that you retreat to a position in which you suspend your belief in them, which is represented as assigning them some probability.

You may also find it useful to contemplate R_0, while taking your knowledge that $R_1, \ldots, R_n$ hold into account. If you suspend your belief in R_0, while taking $R_1, \ldots, R_n$ into account, you will assign R_0 a deliberative probability of $P\left(R_0 \,|\bigcap_{i=1}^{n} R_n\right)$.

To better understand how we might actually assign and interpret deliberative probabilities, let us return to the example from the beginning of this section of promising to go to the movies with Jim. Consider again the reasons R_p, R_ℓ, and R_h introduced above. Perhaps I assume that $P(R_h) = 1$ because I never seriously consider the possibility that Jim is not

[27] Not only do you know that each of $R_0, R_1, \ldots, R_n$ is true, but you also know the status of each of these propositions as a reason for or against action a.

human. Suspending belief about this is not useful for deliberative purposes.

I may assume that $P(R_\ell|R_p) = P(R_\ell)$. In other words, R_ℓ and R_p are independent from the standpoint of deliberative probability. This means that taking R_p to be true for the purpose of deliberation does not cause a change in attitude toward R_ℓ. R_p – that I promised – is a stand-alone reason that conveys no information about R_ℓ – whether Jim loves going to the movies with a friend. Setting $P(R_\ell|R_p) = P(R_\ell)$ amounts to taking the attitude that the promise can be considered as a reason separately from Jim's attitude toward the experience: As a justification, a promise speaks for itself.

Alternatively, I may have a deliberative probability measure such that $P(R_\ell|R_p) > P(R_\ell)$. This could be interpreted as meaning that the reason R_p I am considering is inherently a promise to a friend who (I believe) wants to go to the movies: In deliberation, promising (R_p) inherently conveys information about my friend's desire to go (R_ℓ). In the next section, I will be a little more precise than I have been above and offer an interpretation for what it means for $P(R_0) > \frac{1}{2}$.

4.3 Interpreting Deliberative Probability, Part 2

This section provides a more thorough interpretation of deliberative probability. The analysis is a bit more formal than that of Section 4.2, and relies on the formal properties of weight.

Let us formally define a **weight function** to be a collection $w = (w_a : a \in A)$, where $w_a : \mathscr{G}_+ \to \mathbb{R}$ or all $a \in A$. (Recall that A includes all actions *other than* the default.) *Any* such collection is a weight function. An **expectational** weight function is a weight function that can be derived from that CEV model (see Section 2.2) for some choice of basic value function V and probability measure P via equation (5).

The basic approach I will take here will be to assume that the agent already has a grasp of the weight that various reasons have or would have. We can then use these weights to derive implicit probabilities in the deliberative position. Section 5 below shows that rather than taking value and probability as primitive and deriving weight, we can take weight as primitive and derive value and probability. The derived values and probabilities will be such that it is *as if* the weights were derived from values and probabilities.

Proposition 3 *Let w be an expectational weight function, suppose that both R and $\bar{R}$ belong to $\mathscr{G}_+$, and let a and b belong to A. Then*

1. ***Probability is Implicit in Weight.*** *Suppose that $w_a(R) \neq 0$. Then*

$$P(R) = \frac{w_a\left(\bar{R}\right)}{w_a\left(\bar{R}\right) - w_a(R)} = \frac{\left|w_a\left(\bar{R}\right)\right|}{|w_a(R)| + \left|w_a\left(\bar{R}\right)\right|} = \frac{w_a(R)}{i_a(R)} \tag{25}$$

Consequently,

$$|w_a(R)| > \left|w_a\left(\bar{R}\right)\right| \text{ if and only if } P(R) < \frac{1}{2}. \tag{26}$$

2. ***Opposite Valences of a Reason and Its Negation.*** $w_a(R) > 0$ *if and only if* $w_a\left(\bar{R}\right) < 0$*;* $w_a(R) = 0$ *if and only if* $w_a\left(\bar{R}\right) = 0$*.*

3. ***Constant Weight Ratio of a Reason and Its Negation.*** *If* $w_a\left(\bar{R}\right) \neq 0$*, then* $\frac{w_a(R)}{w_a\left(\bar{R}\right)} = -\frac{1-P(R)}{P(R)}$*; consequently, if* $w_a\left(\bar{R}\right) \neq 0$*, and* $w_b\left(\bar{R}\right) \neq 0$*, then*

$$\frac{w_a(R)}{w_a\left(\bar{R}\right)} = \frac{w_b(R)}{w_b\left(\bar{R}\right)}. \tag{27}$$

The key fact that generates all of the properties in this proposition is the law of iterated expectations. The formula (25), particularly the first equality, is familiar from Bolker-Jeffrey decision theory (Bolker 1967, Jeffrey 1990). Jeffrey expresses probability of a proposition X as the ratio $\frac{des\ \bar{X}}{des\ X - des\ X}$, where *des* X is the desirability of X.[28] Desirability and weight have the same properties because both are conditional expectations.[29]

Equation (25) shows how deliberative probabilities can be derived from weights. The probability of a proposition is the ratio of its weight to its importance. For convenience, I again display Figure 1 here.

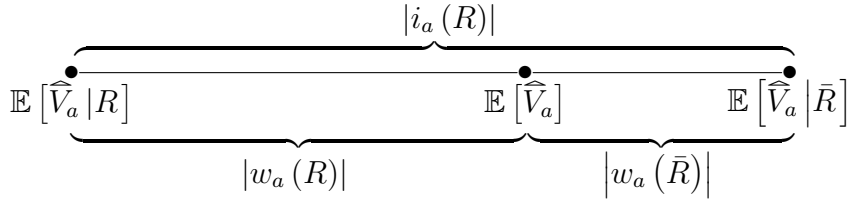

Figure 1: Weight and Importance.

It is instructive to focus on (26), which tells us that a proposition has probability less than $\frac{1}{2}$ precisely if the weight it has for some decision between an alternative a and the default

[28]The above expression for probability in terms of desirability applies when *des* T is normalized to equal 0, where T is a tautology. Also when *des* $T = 0$, Jeffery remarks that *des* X and *des* $\bar{X}$ have opposite signs except when one of these terms is equal to zero, which corresponds to part 2 of Proposition 3.

[29]There are also differences. Desirability concerns value of propositions whereas weight concerns second order differences in value that involve variations in both actions and events as in (6). Substantively, desirability and weight are quite different concepts.

is less than the weight that its negation would have for that decision.[30] Let us get an intuition for why this is so. By the law of iterated expectations $\mathbb{E}\left[\widehat{V}_a\right] = \mathbb{E}\left[\widehat{V}_a \left| R\right.\right] P(R) + \mathbb{E}\left[\widehat{V}_a \left| \bar{R}\right.\right] (1 - P(R))$. In other words, $\mathbb{E}\left[\widehat{V}_a\right]$ is a probability weighted average of $\mathbb{E}\left[\widehat{V}_a \left| R\right.\right]$ and $\mathbb{E}\left[\widehat{V}_a \left| \bar{R}\right.\right]$, with probability weights $P(R)$ and $1 - P(R)$. Suppose that $P(R) < \frac{1}{2}$. Then the probability weight on $\mathbb{E}\left[\widehat{V}_a \left| \bar{R}\right.\right]$ is greater than that on $\mathbb{E}\left[\widehat{V}_a \left| R\right.\right]$. So $\mathbb{E}\left[\widehat{V}_a\right]$ is closer to $\mathbb{E}\left[\widehat{V}_a \left| \bar{R}\right.\right]$ than to $\mathbb{E}\left[\widehat{V}_a \left| R\right.\right]$. This is illustrated by Figure 1 above. As R is less likely than $\bar{R}$, learning that R obtains would move the expected value of $\widehat{V}_a$ more than learning that $\bar{R}$ obtains. On this model, a fact R_0 that is "taken for granted" in the sense that it has a high probability $P(R_0)$ has less weight relative to its negation than a fact R_1 that is informative in the sense that it initially had a low probability $P(R_1)$: That is, $\left|\frac{w_a(R_0)}{w_a(\bar{R}_0)}\right| < \left|\frac{w_a(R_0)}{w_a(\bar{R}_0)}\right|$.

This suggest an alternative way of thinking about deliberative probability. Propositions that have *large* weight (relative to their negations) have a *small* deliberative probability. Intuitively if a proposition R would move our evaluation very much, but its negation would not move our evaluation very much, then that proposition is very informative. Thus, rather than focusing on the probability of a proposition, we can focus on how informative it is. This we can do by turning the probability measure "upside down": Define the **specificity** of a proposition $R \in \mathscr{F}$, $S(R)$ by the relation

$$S(R) = 1 - P(R). \tag{28}$$

Observation 2 *$S : \mathscr{F} \to \mathbb{R}$ is a specificity measure induced by some probability measure on R if and only if:*

$$\forall R_0, R_1, \in \mathscr{F}, \quad S(R_0) + S(R_1) = S(R_0 \cup R_1) + S(R_0 \cap R_1), \tag{29}$$

$$\forall R \in \mathscr{F}, \quad S(R) \geq 0, \quad S(\Omega) = 0, \quad \text{and} \quad S(\emptyset) = 1, \tag{30}$$

$$\forall \{R_i\}_{i=1}^{\infty} \subseteq \mathscr{F}, \quad \lim_{n\to\infty} S\left(\bigcup_{i=1}^{n} R_i\right) = S\left(\bigcup_{i=1}^{\infty} R_i\right). \tag{31}$$

Moreover, if S is a specificity measure, then

$$\forall R_0, R_1 \in \mathscr{F}, \quad R_0 \supseteq R_1 \Rightarrow S(R_0) \leq S(R_1). \tag{32}$$

Condition (29) is an additivity property and (31) is a continuity property. A probability measure satisfies all of the properties (29)-(31) except that $P(\emptyset) = 0$ and $P(\Omega) = 1$ whereas $S(\emptyset) = 1$ and $S(\Omega) = 0$; that is, the values assigned to $\emptyset$ and to Ω are exchanged. Likewise

[30] Observe that if $|w_a(R)| < |w_a(\bar{R})|$ for some a, then for all b, if $w_b(R) \neq 0$, then $|w_b(R)| < |w_b(\bar{R})|$.

the monotonicity condition is reversed: more inclusive propositions are more probable, but less inclusive propositions are more specific.

Since probability and specificity are interdefinable, we can take either as primitive. It may be more intuitive to think in terms of specificity. It will help to consider some examples. Consider the example of promising you that I would go to the movies from Section 2.2.6. It is intuitive that the proposition that I promised you would be more informative than the proposition that I did not promise you. This accounts for the greater weight that promising has as a reason for going than not promising would as a reason for not going.

Let us consider a normative proposition such as the proposition R that it is wrong to kill. The proposition R can serve as a reason and it can have weight, and yet it may be difficult to grasp what it means to assign such a normative proposition a probability.[31] First, observe that if we can assign this reason a weight, then its deliberative probability can be viewed as implicit in its weight via (25). Second, it is intuitive to say that the proposition that it is wrong to kill, assuming that it is true, is much more informative than would be its negation that it is not wrong to kill. We have seen that informativeness or specificity is sufficient to determine deliberative probability. To make this vivid, imagine that you know nothing about morality. Then, to learn that it is wrong to kill would dramatically change your worldview; it would be very informative. Perhaps it would also be informative to learn that it is not wrong to kill, but much less so. Accordingly that it is wrong to kill is a very weighty consideration when it is relevant, whereas, that it is not wrong to kill, were it true, would not argue for much.

Let us revisit Figure 1. Rather than thinking of the importance $i_a(R)$ as the sum of weights $w_a(R)$ and $-w_a\left(\bar{R}\right)$ (see Definition 3), we can think of the weight of R as a share of the importance of R: that is, that there exists a number $\lambda_a(R)$ between 0 and 1 such that:

$$w_a(R) = \lambda_a(R)\, i_a(R).$$

We refer to $\lambda_a(R)$ as R's **share** of the importance. Likewise $\lambda_a\left(\bar{R}\right) = 1 - \lambda_a(R)$ is $\bar{R}$'s share of the importance. $\lambda_a(R)$ represents the proportion of the movement from $\bar{R}$ to R which occurs in the second half, that is, in the movement from the state of ignorance to knowledge of R.

[31] To articulate just one aspect of the difficulty, the proposition that it is wrong to kill does not appear to be a contingent fact that may be true in some states of the world and false in others. In order to assign probabilities (not equal to zero or one) to begin with, we must however be able to imagine "possible worlds" in which it is true that it is wrong to kill and others in which it is false. Observe that we can ask the question whether it is wrong to kill. So it seems that there is the *conceptual* possibility that it is wrong to kill and the conceptual possibility that it is not wrong to kill. If the model is to apply to normative propositions, then such conceptual possibilities must be modeled as possible states of the world; for instance, there is a conceptually possible world in it is wrong to kill and others in which it is not.

A consequence of the CEV model is that the share $\lambda_a(R)$ does not depend on the alternative a that is opposed to the default. Indeed the share of R is its specificity $S(R) = \lambda_a(R)$ for all $a \in A$. This does not mean that the *weight* of a reason is constant across actions. The weight of a reason R $w_a(R)$ – its sign and magnitude – and its weight relative to another reason R', $\frac{w_a(R)}{w_a(R')}$ can vary across actions a because the importance of the reason $i_a(R)$ can vary across actions, but the weight of a reason for an action relative to the weight that its negation would have for that action is constant across actions. (The one exception is in the case that the reason is irrelevant to the action so that $w_a(R) = 0$.) This is expressed by (27) and is illustrated by Figure 2.

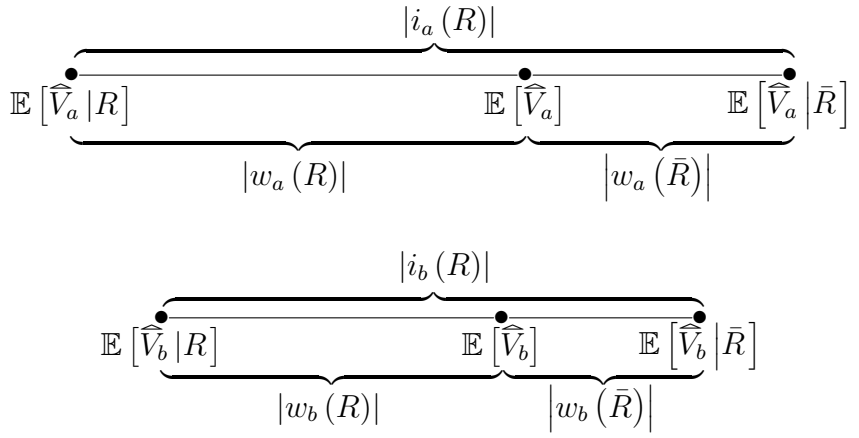

Figure 2: The importance of R varies across actions but the share of R is constant.

More generally than (27), the CEV model actually implies that for any four alternatives a, b, c, and e, $\frac{w_{a,b}(R)}{w_{a,b}\left(\bar{R}\right)} = \frac{w_{c,e}(R)}{w_{c,e}\left(\bar{R}\right)}$.[32]

Is it plausible that the share of a reason R is constant across decisions? If we think about propositions such as *It is wrong to harm people* or *Ann lacks basic resources*, it is plausible that these reasons are much more significant than their negations would be regardless of the decision under consideration. This lends some intuitive plausibility to the constant share hypothesis. It is important to note that the constant shares condition holds as we vary the two alternatives a and b, but the background information, or other reasons already conditioned on much be held fixed for the constant shares condition to hold. If $\mathscr{A}$ and $\mathscr{B}$ are two distinct sets of reasons, the CEV model does *not* imply that $\frac{w_a(R|\mathscr{A})}{w_a\left(\bar{R}|\mathscr{A}\right)} = \frac{w_b(R|\mathscr{B})}{w_b\left(\bar{R}|\mathscr{B}\right)}$.

[32]A further prediction is that if R_0 is more informative than R_1, in the sense that $R_0 \subseteq R_1$, then R_0 will be weightier relative to its negation than R_1 would be: $\frac{w_a(R_0)}{w_a\left(\bar{R}_0\right)} > \frac{w_a(R_1)}{w_a\left(\bar{R}_1\right)}$. This does not mean that R_0 is weightier than R_1: It is *not* implied that $w_a(R_0) > w_a(R_1)$.

I conclude this section with an example for which the constant shares condition is problematic. The fact that I promised you I would go is a reason to go, and it is also a reason to apologize if I don't go. My promise may be more important to the decision of whether I go or don't go than to the decision of whether I don't go but apologize or don't go and don't apologize; the model does not predict that my promise will be equally important to both.

The theory does predict that if the fact that I promised is a strong reason to go but the fact that I did not promise would have been a *very* weak reason not to go, then the fact that I promised would have been a much stronger reason to apologize and not go rather than not apologize and not go than the fact that I did not promise would have been not to apologize and not go as opposed to apologize and go. This may be so. But note that the fact that I did not promise is at best a very weak reason not to go. On the other hand the fact that I did not promise may be quite a significant reason not to apologize not go rather than to apologize and not go. Let a, b, c and d be respectively, go, don't go, don't go and apologize, and don't go and don't apologize. (Action b may include any assumption about apology that one wishes.) Then the above considerations suggest that $\left|\frac{w_{a,b}(R)}{w_{a,b}(\bar{R})}\right| > \left|\frac{w_{c,e}(R)}{w_{c,e}(\bar{R})}\right|$, contrary to the theory.

Note that in the last paragraph I used the phrase "apologize and not go rather than not apologize and not go" rather than "apologize if I don't go" because the latter might amount to conditioning on additional information, which would nullify the constant shares assumption.[33] This makes our intuitions about the example, a little less clear. Still, I think the example is problematic even given this clarification. The theory will have to address examples such as this.

4.4 Integrating Deliberative and Actual Uncertainty

In many decision situations, we know the truth of some propositions that are relevant to our decisions while being uncertain about others. We must therefore weigh the reasons that we are aware of while considering possibilities of which we are unaware, but which are relevant to our decisions. This suggests a combination of deliberative and actual probability. Indeed, model encourages thinking in this way. For example, when Proposition 2 analyzes maximization of expected value conditional on the reasons that we have, it suggests that there may be other relevant facts out there that we do not know about.

Integrating these different types of probability is complicated. One approach would be to interpret the probability measure over the field of facts we know as deliberative probability and the probabilities over uncertain events conditional on the facts we know as actual

[33]In the paragraph before that, when introducing the example, I used "apologize if I don't go" because it is more intuitive.

probabilities. However, this may be an unstable mixture, especially when one considers the dynamics of updating beliefs. Another approach would be to rigorously separate these two kinds of probability. We imagine that we know all the relevant facts and derive the relevant weights by updating to this situation from the deliberative position. Then we step back and account for our actual ignorance and calculate expected weights based on our actual uncertainty. Future work will have to consider these alternatives more closely.

5 Equivalence of Weight and Value

This section establishes an equivalence between weight and value. It shows that we can either take weight or value as primitive and derive the other. Thus, the two alternatives that are open to us are:

1. **Weight Derived From Value.** We can start with value as primitive, and derive the notion of weight from value. Starting with value really means starting with value *and* probability.

2. **Value Derived From Weight.** Alternatively, we can start with weight as primitive. If we impose certain axioms on weight, we can then derive both value and probability from weight.

This section establishes the equivalence. This equivalence means that the CEV model need not to be bundled with the thesis that value is more fundamental with weight. The CEV model rather establishes a structural relation and a translation manual between value and weight, provided that weight satisfies certain axioms. We can then argue on philosophical grounds about which notion, if either, should be treated as primitive. Or we may prefer to remain agnostic.

Let us assume for simplicity that $\mathscr{F} = \mathscr{G}$, that is, that every proposition in $\mathscr{F}$ is a potential reason. I modify the definition a weight function slightly. I call a collection $w = (w_a : a \in A)$, where $w_a : \mathscr{F} \to \mathbb{R}$ and $w_a(\emptyset) \neq 0$ for all $a \in A$ a **weight function**. For any action $a \in A$ and proposition R in $\mathscr{F}_a$, as above, $w_a(R)$ is the weight of R for or against a. Recall that A includes all actions *other than* the default. So no weight is assigned to the default.[34]

[34]There are two slight modifications here in the definition of a weight function, in comparison to the definition in Section 4.3. First, the domain of each functon w_a is $\mathscr{F}$ rather than $\mathscr{F}_+$. The reason for this is that, here I will derive a value function from a probability measure and a weight function. Therefore, I do not want to presume which propositions have probability zero ahead of time. The second difference is that I have assumed that $w_a(\emptyset) \neq \emptyset$. Since $\emptyset$ corresponds to the inconsistent proposition, we are not

In the CEV model, $w_a(R) = \mathbb{E}\left[\widehat{V}_a \,|R\right] - \mathbb{E}\left[\widehat{V}_a\right]$. To simplify the discussion, let us assume for the moment that $\mathbb{E}\left[\widehat{V}_a\right] = 0$ for all $a \in A$, although this will not be assumed for the formal results below. Nothing of technical importance is affected by this simplification. Under the simplification $w_a(R) = \mathbb{E}\left[\widehat{V}_a \,|R\right]$. It is useful at this point to indicate that the expectation is with respect to probability measure P, and so I write $w_a(R) = \mathbb{E}_P\left[\widehat{V}_a \,|R\right]$.

We would like to impose certain axioms on w such that whenever these axioms are satisfied there exist V and P such that $w_a(R) = \mathbb{E}_P\left[\widehat{V}_a \,|R\right]$ for all $a \in A$ and $R \in \mathscr{F}$ with $P(R) > 0$. Let us decompose this into two steps. We want to show that there exist a value function V and a collection of probability measures $(P_a : a \in A)$ such that

1. (**Conditional Expectation**) $w_a(R) = \mathbb{E}_{P_a}\left[\widehat{V}_a \,|R\right]$ for all $R \in \mathscr{F}$ and $a \in A$, and
2. (**Common Probability**) $P_a = P_b$ for all $a, b \in A$.

Bolker (1966) axiomatized the conditions under a function ϕ taking events R as inputs has the structure of conditional expectation.[35] This answers the question: For what functions ϕ with events as inputs does there exists a random variable X and probability measure $\tilde{P}$ such that $\phi(R) = \mathbb{E}_{\tilde{P}}[X \,|R]$? Bolker presented two axioms which are sufficient for ϕ to have the structure of conditional expectation. One of Bolker's axioms is called averaging.

> **Averaging:** for all R_1, R_2 with $R_1 \cap R_2 = \emptyset$,
> $\phi(R_1) \leq \phi(R_2) \Rightarrow \phi(R_1) \leq \phi(R_1 \cup R_2) \leq \phi(R_2)$, and

The other is called impartiality.[36,37] Bolker was working in a slightly different framework than the one I am working with here.[38] His axioms are only necessary but not sufficient

really interested in the weight of $\emptyset$. It is however technically convenient to assign this proposition a weight. Any nonzero number will do. Unlike the expression, $w_a(R)$ for nonempty R, which represents weight in its intuitive sense, $w_a(\emptyset)$ is not intended to be given any intuitive meaning.

[35]Formally, Bolker axiomatized functions ϕ of the form $\phi(R) = \frac{M(R)}{P(R)}$, where M is a signed measure and P is a probability measure. The connection to the structure of conditional expectation is for a random variable X, and event R with $P(R) > 0$, $\mathbb{E}_P[X \,|R] = \frac{M(R)}{P(R)}$ where $M(R) = \int_R X dP$. By the Radon-Nikodym theorem a signed measure $M(R)$ is always of the form $\int_R X dP$ when M is absolutely continuous with respect to P.

[36]The impartiality axiom says that for all R_1, R_2, and R_3 which are pairwise disjoint and are such that $\phi(R_1) = \phi(R_2)$ but $\phi(R_3) \neq \phi(R_2)$, if $\phi(R_1 \cup R_3) = \phi(R_2 \cup R_3)$, then for all R_4 which are pairwise disjoint from $R_1 \cup R_2$, $\phi(R_1 \cup R_4) = \phi(R_2 \cup R_4)$.

[37]Variants on the impartiality and averaging that appeal to an ordering on propositions rather than a function ϕ on propositions form the basis for Bolker-Jeffrey decision theory.

[38]In particular, for the result in question, Bolker worked with a complete atom free Boolean algebra, whereas I work with a σ-algebra.

in my framework for the expectational structure criterion. These differences are primarily technical rather than substantive.

I will develop a different axiomatization for the conditional expectation property and then impose an additional assumption that imposes common probability. Let me sketch the main idea. Define $M_a(R) = \int_R \widehat{V}_a dP_a$. Our aim is that w_a will be of the form $w_a(R) = \mathbb{E}_P\left[\widehat{V}_a | R\right]$. Equivalently, w_a will be of the form $w_a(R) = \frac{M_a(R)}{P_a(R)}$. Observe that M_a is a signed measure and P_a is a probability measure. As we saw in Section 4.3 – see specifically (25) – probability P_a can be defined in terms of w_a. Since $M_a(R) = w_a(R) P_a(R)$, it now also follows that M_a can be defined in terms of weight. So we can simply impose as our basic axioms that P_a is a probability measure and M_a is a signed measure. This gets us the conditional expectation property. To get the common probability property, we then simply assume that $P_a = P_b$, as this equality can be spelled out in terms of weight. Equivalently, we assume the constant weight ratio property (27).

The one obstacle to the above approach is that P_a is not defined when $w_a(R) = 0 = w_a\left(\bar{R}\right)$. I now develop the approach a little more fully to overcome this issue.

Consider a weight function $w = (w_a : a \in A)$. For each $a \in A$, let

$$\mathscr{F}_a^w := \left\{R \in \mathscr{F} : w_a(R) \neq w_a\left(\bar{R}\right)\right\}, \tag{33}$$

$\mathscr{F}_a^w$ is the set on which implicit probabilities are well defined if we attempt to define them according to (25). For each $a \in A$, define $P_a^w : \mathscr{F}_a^w \to \mathbb{R}$ and $M_a^w : \mathscr{F} \to \mathbb{R}$ by

$$P_a^w(R) = \frac{w_a\left(\bar{R}\right)}{w_a\left(\bar{R}\right) - w_a(R)} \quad \forall R \in \mathscr{F}_a^w, \tag{34}$$

$$M_a^w(R) = \begin{cases} \frac{w_a(R) w_a\left(\bar{R}\right)}{w_a\left(\bar{R}\right) - w_a(R)} & \text{if } R \in \mathscr{F}_a^w, \\ 0 & \text{if } R \notin \mathscr{F}_a^w, \end{cases} \quad \forall R \in \mathscr{F}. \tag{35}$$

These equations express how P_a and M_a are derived from w as discussed above. I have indexed P_a and M_a by the superscript w to indicate that they are derived from w.

I now impose the following conditions:

Continuity: $\forall a \in A, \forall \{R_i\}_{i=1}^{\infty} \subseteq \mathscr{F}, w_a\left(\bigcup_{i=1}^n R_i\right) = \lim_{n\to\infty} w_a\left(\bigcup_{i=1}^n R_i\right)$.

Valence: $\forall a \in A, \forall R \in \mathscr{F}, w_a(R) = w_a\left(\bar{R}\right) \Rightarrow w_a(R) = 0$.

Implicit Probability: $\forall a \in A, \forall R_1, R_2, R_3, R_4 \in \mathscr{F}_a^w$,

$$\begin{aligned}(R_1 \cap R_2 = R_3 \cap R_4 \text{ and } R_1 \cup R_2 = R_3 \cup R_4) \\ \Rightarrow P_a^w(R_1) + P_a^w(R_2) = P_a^w(R_3) + P_a^w(R_4).\end{aligned}$$

Implicit Measure: $\forall a \in A, \forall R_1, R_2 \in \mathscr{F}$,

$$R_1 \cap R_2 = \emptyset \Rightarrow M_a^w(R_1) + M_a^w(R_2) = M_a^w(R_1 \cup R_2).$$

Common Ratios: $\forall a, b \in A, \forall R \in \mathscr{F}, \left(w_a\left(\bar{R}\right) \neq 0 \text{ and } w_b\left(\bar{R}\right) \neq 0\right) \Rightarrow$

$$\frac{w_a(R)}{w_a\left(\bar{R}\right)} = \frac{w_b(R)}{w_b\left(\bar{R}\right)}. \tag{36}$$

Implicit Measure and Continuity imply that M_a^w is a signed measure. I believe that Implicit Probability, Continuity and Implicit Measure jointly imply that P_a^w is extendible to a probability measure on $\mathscr{F}$. I have a proof sketch for this last statement, but have not yet verified all details (see Remark 1). Common Ratios is equivalent to the claim that $P_a^w(R) = P_b^w(R)$ for all $R \in \mathscr{F}_a^w$.

In Section 4.3, an expectational weight function was defined as one that is generated by the CEV model. In this section, I would like to define an expectational model axiomatically rather than in terms of the CEV model. So I reset the meaning of the term "expectational" according to the Definition 7 below.

Definition 7 *A weight function w is **expectational** if it satisfies Continuity, Valence, Implicit Measure, Implicit Probability, and Common Ratios.*

Proposition 5* *Let w be a weight function. Then the following conditions are equivalent.*

1. *w is expectational.*
2. *There exists a probability measure P on $(\Omega, \mathscr{F})$ and a collection $(V_a : a \in A_0)$ of measurable real-valued functions on $(\Omega, \mathscr{F})$ such that:*

$$\forall R \in \mathscr{F}, \quad P(R) > 0 \Rightarrow w_a(R) = \mathbb{E}_P[V_a - V_d \,|R] - \mathbb{E}_P[V_a - V_d]$$

Suppose that w is expectational and there exists $b \in A$ and $R \in \mathscr{F}$ such that both $w_b(R) \neq 0$ and $w_b\left(\bar{R}\right) \neq 0$. Then, in 2, P is uniquely determined and for any choice of numbers $(\delta_a : a \in A)$ – these δ_a's are default advantages (see Section 3.3) – 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 $a \in A$, $\widehat{V}_a = V_a - V_d$ is uniquely determined up to a set of P-measure zero, while V_d can be chosen freely. Thus by setting $\delta_a = 0$, for all $a \in A$, we may always impose initial neutrality consistently with w.

Remark 1 *Proposition 5 is marked with a star because I have not yet verified all details of the proof. I have a detailed proof sketch and have verified many of the details but not all. In the remainder of this section I will write as if the proposition were established.*

The proposition spells out the what must be assumed about weight in order that a value function and deliberative probability measure can be derived from weight. We saw in Section 4.3 how probability P could be derived from weight, namely via (25). How can value be derived from weight? Comparative value can be derived from the Radon-Nikodym derivative of $w_a(R) P(R)$ with respect to $P(R)$. For any δ, there is an (essentially) unique solution $\widehat{V}_a$ to:

$$\widehat{V}_a + \delta_a = \frac{dM_a^w}{dP}$$
$$\mathbb{E}\left[\widehat{V}_a\right] = \delta_a$$

To understand this, observe that when we don't assume that $\mathbb{E}\left[\widehat{V}_a\right] \neq 0$, then in general we have $M_a^w(R) = w_a(R) P(R) = \int_R \widehat{V}_a dP + \mathbb{E}\left[\widehat{V}_a\right] P(R)$. If we assume initial neutrality, then comparative value is completely determined by weight. In general, if we have the vector $\delta = (\delta_a : a \in A)$ of default advantages and and an expectational weight function, then comparative value $\widehat{V}_a = V_a - V_d$ is completely determined. However V_a and V_d are not determined separately.

It is important to separate two claims: (i) Value and deliberative probability can be derived from weight. (ii) The axioms on weight under which this is so are correspond to intuitive properties of weight. Proposition 5 spells out the details of (i). In the absence of Proposition 5, it would still be clear that there is some class of weight functions such that if weight is in that class, value and probability are implicitly determined. Proposition 5 spells out axiomatically what that class is. Proposition 5 does not perform as well on (ii). In particular, it is not intuitively clear that weight should satisfy Implicit Probability and Implicit Measure; at least, I do not have a direct argument for this at this time. An alternative would be to substitute for Implicit Probability and Implicit Weight, Bolker's Averaging and Impartiality Axioms applied to weight. For example, applied to weight, the averaging axiom says that if $R_1 \cap R_2 = \emptyset$, if $w_a(R_1) \leq w_a(R_2)$, then $w_a(R_1) \leq w_a(R_1 \cup R_2) \leq w_a(R_2)$. This is a plausible property of weight, as is impartiality (see footnote 36). To take this approach would require strengthening other technical assumptions. For example, we have not assumed that $\mathscr{F}$ is atom free; for example, $\mathscr{F}$ is allowed to be finite in the above result. Note that, given the appropriate technical modifications, in addition to Averaging and Impartiality, Common Ratios would have to be assumed. Finally observe that an *indirect* argument for Implicit Probability and Implicit Measure is that under the appropriate technical modifications, Averaging and Impartiality would imply Implicit Probability and Implicit Measure.

Finally, I mention that in the same way that Bolker-Jeffrey decision theory derives desirability and probability functions from a preference ordering on propositions, it may be possible to derive weights, values and deliberative probabilities from a family of strength orderings on reasons.[39]

6 Conclusion

This paper has presented a formal model of reasons. I have linked comparative value to the weight of reasons. The weight of a reason is analyzed in terms of the way that considering the reason alters the comparison between an action and a default. The theory allows one to study how reasons interact and characterizes independent reasons. The theory allows one to take either value or weight as primitive and derive the other.

Much remains to be done. One task is to use the model further analyze the two types of conflict among reasons described in the introduction. Another is to apply the theory to the aggregation of reasons of different individuals, or to help us understand what we should do if aggregation is not appropriate.

A Proofs

The proofs will be inserted in a future draft.

B Independence Versus Strong Independence

In this appendix, I present an example that shows that strong a-independence does not imply independence. Consider the following tables:

$\widehat{\widehat{V}}_a$	R_2	$\sim R_2$
R_1	0	-1
$\sim R_1$	1	0

P	R_2	$\sim R_2$
R_1	$1/3$	$1/6$
$\sim R_1$	$1/6$	$1/3$

In both tables, I have used the notation $\sim R$ instead of $\bar{R}$ for the complement of R to enhance readability. The table on the left gives the value of a $\widehat{\widehat{V}}_a$ as a function of the state ω. I assume that $\widehat{\widehat{V}}_a$ depends only on whether each of the events R_1 and R_2 obtain or fail to obtain. The

[39] This strength ordering would have to compare the strength of reason R_1 for action a to the strength of R_2 for b.

table on the right gives the joint probability of R_1 and R_2 and their complements. Using these tables, we calculate the following (conditional) expectations:

$$\begin{aligned}
\mathbb{E}\left[\widehat{V}_a\right] &= 0, \\
\mathbb{E}\left[\widehat{V}_a \,|R_1\right] &= -\frac{1}{3}, \\
\mathbb{E}\left[\widehat{V}_a \,|R_2\right] &= \frac{1}{3}, \\
\mathbb{E}\left[\widehat{V}_a \,|R_1 \cap R_2\right] &= 0, \\
\mathbb{E}\left[\widehat{V}_a \,\middle|\bar{R}_1\right] &= \frac{1}{3}, \\
\mathbb{E}\left[\widehat{V}_a \,\middle|\bar{R}_1 \cap R_2\right] &= 1.
\end{aligned}$$

So

$$\begin{aligned}
w_a\left(R_1\right) &= \mathbb{E}\left[\widehat{V}_a \,|R_1\right] - \mathbb{E}\left[\widehat{V}_a\right] = -\frac{1}{3} = \mathbb{E}\left[\widehat{V}_a \,|R_1 \cap R_2\right] - \mathbb{E}\left[\widehat{V}_a \,|R_2\right] = w_a\left(R_1 \,|R_2\right), \\
w_a\left(R_2\right) &= \mathbb{E}\left[\widehat{V}_a \,|R_2\right] - \mathbb{E}\left[\widehat{V}_a\right] = \frac{1}{3} = \mathbb{E}\left[\widehat{V}_a \,|R_1 \cap R_2\right] - \mathbb{E}\left[\widehat{V}_a \,|R_1\right] = w_a\left(R_2 \,|R_1\right).
\end{aligned}$$

So $\{R_1, R_2\}$ is a-independent. On the other hand,

$$w_a\left(R_2 \,\middle|\bar{R}_1\right) = \mathbb{E}\left[\widehat{V}_a \,\middle|\bar{R}_1 \cap R_2\right] - \mathbb{E}\left[\widehat{V}_a \,\middle|\bar{R}_1\right] = \frac{2}{3} \neq \frac{1}{3} = w_a\left(R_2\right).$$

So $\{R_1, R_2\}$ is not strongly a-independent. It follows that a-independence does not imply strong a-independence.

C An Example: Utilitarianism

This section presents the example briefly discussed in Section 3.5. Specifically, this section examines the weight of reasons from the standpoint of a utilitarian.

There are two agents, Ann and Bob. Given any action a, value given that a is taken is represented as:

$$V_a = U_a^A + U_a^B. \tag{37}$$

U_a^A and U_a^B are utilities of Ann and Bob that depend on the state of the world conditional on action a being taken. Formally, U_a^A and U_b^A are random variables. That value is given by the sum of utilities reflects the utilitarian values mentioned above.

Suppose that you are considering a project a that will help Ann and harm Bob. Let R_A be the proposition that the the project would help Ann. Let R_B be the proposition that the project would harm Bob.

Assume that the default d is a status quo that involves doing nothing. Specifically $U_d^A \equiv 0$ and $U_d^B \equiv 0$, or, in words, the default leads to a utility of 0 in every state of the world.

Suppose that for the purpose of deliberation you treat R_A and R_B as being independent of one another (i.e., in deliberating about the project, when we consider the force of the fact that it will help Ann, we do not take into account the effect on Bob). Formally:

1. U_a^B and R_A are mutually independent both unconditionally and conditional on R_B and on $\bar{R}_B$.

2. U_a^A and R_B are mutually independent both unconditionally and conditional on R_A and on $\bar{R}_A$.

Then Proposition 2 implies that R_A and R_B are independent reasons for/against the project.

Now suppose that the project has two sorts of impacts: R_W means that the project would have a positive impact on agents' wealth. R_H means that the project would have a negative impact on agents' health. For Ann, the wealth effect outweighs the health effect, which is why it helps her. For Bob, the health effect outweighs the wealth effect, which is why it harms him. Specifically, the relation between, wealth, health and utility is expressed as follows

$$U_a^A = u^A(W_a^A, H_a^A) \text{ and } U_a^B = u^B(W_a^B, H_a^B).$$

Here W_a^A and H_a^A are real valued random variables that encode Ann's level of wealth and health at each state of the world, and u^A is a utility function that maps Ann's wealth and health to her utility level. I assume that these two variables, wealth and health, determined Ann's utility. W_a^B, H_a^B, and u^B are defined similarly, but pertain to Bob rather than Ann.

Say that utility is **separable in wealth and health** if there exist subutility functions u_w^A, u_h^A, u_w^B and u_h^B such that

$$u^A(W_a^A, H_a^A) = u_w^A(W_a^A) + u_h^A(H_a^A),$$
$$u^B(W_a^B, H_a^B) = u_w^B(W_a^B) + u_h^B(H_a^B).$$

u_w^A and u_h^A are, respectively, Ann's utility of wealth and utility of health functions, and u_w^B and u_h^B are defined similarly for Bob. Then,

$$V_a = u_w(W_a) + u_h(H_a)$$

where, $W_a = (W_a^A, W_a^B)$ and $H_a = (H_a^A, H_a^B)$, $u_w(W_a) = u_w^A(W_a^A) + u_w^B(W_a^B)$ $u_h(H_a) = u_h^A(H_a^A) + u_h^B(H_a^B)$. In other words, rather than expressing the value of action a as the sum of Ann and Bob's utilities as in (37), we can rewrite utility as the sum of aggregate wealth utility and aggregate health utility.

Let R_W be the proposition that the project would have a positive impact on agents' wealth and R_H the proposition that the project would have a negative impact on agents' health. If

1. utility is separable in wealth and health,

2. (i) $u_w(W_a)$ and R_H are independent both unconditionally and conditional on R_W and on $\bar{R}_W$, and

   (ii) $u_h(H_a)$ and R_W are independent both unconditionally and conditional on R_H and on $\bar{R}_H$,

Proposition 2 implies that R_W and R_H are independent reasons. If either condition 1 or 2 fails, then R_W and R_H will generally not be independent reasons. 2 might fail because health and wealth may be causally interrelated and we may want to take this into account in deliberation.

An action (e.g., our project) may have multiple rationales. There may be multiple ways of cutting up the world. Some (e.g., R_A and R_B) may be independent – especially if they are in terms of events that concern fundamental ethical categories – in this case the utilitarian takes Ann and Bob's utilities to be fundamental ethical values out of which overall value is built. Others (e.g., R_W and R_H) may or may not be independent.

An interesting question that arises here is: Should we prefer independent reasons (R_A and R_B) to non-independent reasons (R_W and R_H) when both are available?

Note that in some cases, the competing rationales may each be *complete* (each is redundant given the other) and they may or may not be *equally informative* (they may cut the world into finer or courser slices). I illustrate this through a slight modification of the example.

I modify the example slightly. R_A remains the proposition that the project would *help* Ann, but R_B now becomes the proposition that the project would *help* Bob. Let R_S be the proposition that the project has the same effect on both: that is, either the project helps both Ann and Bob or the project harms both Ann or Bob. Suppose that if the project helps an agent that leads to an additional utility of 1 relative to the default, and when it harms an agent, that leads to a loss of a utility of 1, relative to the default. We can represent the possibilities in two equivalent ways in two different tables.

V_a	R_B	$\sim R_B$
R_A	2	0
$\sim R_A$	0	-2

V_a	R_S	$\sim R_S$
R_A	2	0
$\sim R_A$	-2	0

Figure 5: Two Ways of Cutting Up The Possibilities

Here, I use the notation $\sim R$ instead of $\bar{R}$. Assume that each of the four possibilities in each of the tables is equiprobable. Observe that the sets of reasons $\mathscr{R}^0 = \{R_A, R_B\}$ and $\mathscr{R}^1 = \{R_A, R_S\}$ are *logically equivalent* not only in that $R_A \cap R_B = R_A \cap R_S$ but in that they generate the same field (i.e, set of events constructible via intersection and complementation), so they are in a sense equally informative. Yet only $\mathscr{R}^0$ is independent:

$$w_a(R_A|R_B) = 2 - \left[\frac{1}{2}2 + \frac{1}{2}0\right] = 1 = \left[\frac{1}{2}2 + \frac{1}{2}0\right] - 0 = w_a(R_A)$$
$$w_a(R_A|R_S) = 2 - 0 = 2 > 1 = \left[\frac{1}{2}2 + \frac{1}{2}0\right] - 0 = w(R_A)$$

While the arguments in terms of the reasons $\mathscr{R}^0$ and $\mathscr{R}^1$ are in logically equivalent, there is a case to be made that the argument in terms of $\mathscr{R}^0$ is more basic, as reflected by the independence of the reasons to which it appeals.

References

Alvarez, M. (2016), 'Reasons for action: justification, motivation, explanation', *The Stanford Encyclopedia of Philosophy* . http://plato.stanford.edu/archives/sum2016/entries/reasons-just-vs-expl/.

Arrow, K. (1951), 'Social choice and individual values'.

Billingsley, P. (1995), *Probability and measure*, John Wiley & Sons.

Bolker, E. D. (1966), 'Functions resembling quotients of measures', *Transactions of the American Mathematical Society* **124**(2), 292–312.

Bolker, E. D. (1967), 'A simultaneous axiomatization of utility and subjective probability', *Philosophy of Science* pp. 333–340.

Broome, J. (1991), *Weighing goods*, Blackwell Publishers.

Broome, J. (2013), *Rationality through reasoning*, John Wiley & Sons.

Chang, R. (2016), Comparativism: The grounds of rational choice, *in* E. Lord and B. Maguire, eds, 'Weighing Reasons', Oxford University Press.

Dasgupta, P. (2001), *Human well-being and the natural environment*, Oxford University Press.

Dietrich, F. and List, C. (2007), 'Arrow's theorem in judgment aggregation', *Social Choice and Welfare* **29**(1), 19–33.

Dietrich, F. and List, C. (2013), 'A reason-based theory of rational choice', *Nous* **47**(1), 104–134.

Dietrich, F. and List, C. (2016), 'Reason-based choice and context-dependence: An explanatory framework', *Economics and Philosophy* **32**(02), 175–229.

Jeffrey, R. C. (1990), *The logic of decision*, University of Chicago Press.

Kearns, S. and Star, D. (2009), 'Reasons as evidence', *Oxford studies in metaethics* **4**, 215–42.

List, C. and Pettit, P. (2002), 'Aggregating sets of judgments: An impossibility result', *Economics and Philosophy* **18**(01), 89–110.

List, C. and Polak, B. (2010), 'Introduction to judgment aggregation', *Journal of economic theory* **145**(2), 441–466.

Lord, E. and Maguire, B., eds (2016), *Weighing Reasons*, Oxford University Press.

Mongin, P. (1997), 'Spurious unanimity and the pareto principle'. Working Paper, Theorie Economique, Modelisation et Applications, Université de Cergy-Pontoise and Centre National de Recherche Scientifique.

Raz, J. (1999), *Practical reason and norms*, Oxford University Press.

Scanlon, T. (1998), *What we owe to each other*, Harvard University Press.