The Logic of Reasons

Shyam Nair and John Horty

*** Draft ***

Version of: August 7, 2014

Contents

1 Introduction

Reasons figure large in our ordinary talk of deliberating about or justifying actions or conclusions. Suppose, for example, you want to convince a friend to dine with you at Obelisk tonight. Typically, you will offer reasons—there is a new chef, the reviews have been excellent. Or suppose you want to explain why you believe raccoons have been in the back yard. You will offer your evidence, again, typically, in the form of reasons—the garbage was broken into, those tracks look like raccoon prints.

In spite of their prevalence in our ordinary talk, however, reasons appear to play no role at all in the two predominant formal theories of deliberation and justification: standard logic, and decision theory. In logic, for example, we learn how to establish that premises imply conclusions, sometimes in accord with formal rules of inference, sometimes through a model-theoretic definition. In decision theory, we learn how to rank proposed actions by expected utility, defined in terms of the possible outcomes of those actions, their probabilities, and utilities. Neither of these well-established theories involves any appeal to our ordinary concept of a reason.

Perhaps this is as it should be—perhaps the concept of a reason is nothing but an artifact of folk psychology that should be eliminated from serious scientific discourse. Even so, it may still be useful to map out what logic there is underlying our talk of reasons: the architecture of our folk psychology is not without interest. But it is also possible that the concept of a reason is not eliminable at all, but central to an understanding of ourselves and the world we live in. This idea can be found at times in epistemology, and especially in ethics, where many writers take the notion of a reason as basic to the normative realm, a fundamental concept in terms of which other normative notions can then be analyzed. If this is correct, then understanding the logic of reasons is not just interesting, but crucial.

What would such a logic look like? Just as ordinary logics show us how premises interact to yield conclusions, a logic of reasons would have to show us, at the very least, how reasons interact to support the outcomes they do. Suppose, for example, that you have arranged to meet Emma for lunch. You therefore have a reason to do so, and so it seems that you ought to. But why, exactly? How does the reason generate the ought? Suppose that, on the way to meet Emma for lunch, you encounter your friend Max drowning in a canal. The duties of friendship, and perhaps other duties, now provide you with a reason to rescue Max—which would prevent you, however, from meeting Emma for lunch. It is natural conclude that what you ought to do, in this situation, is rescue Max, rather than meet Emma. But again, why? How do these two reasons interact to support one outcome, rather than the other?

Or consider an epistemic case. Suppose Eric tells you that Tweety is a bird. Trusting Eric as you do, you now have a reason for the conclusion that Tweety is a bird, which itself forms a reason for the further conclusion that Tweety can fly—since birds, typically, can fly. But suppose also that Rachael, who is equally trustworthy, tells you that Tweety is a penguin. You now have a reason for the conclusion that Tweety is a penguin, which forms a reason for the further conclusion that Tweety cannot fly—since penguins cannot fly. Here, it is natural to accept both of the initial conclusions—that Tweety is both a bird and a penguin—but to reject the conclusion that Tweety can fly in favor of the conclusion that he cannot. But why? How do these various reasons interact to support the conclusions that Tweety is both a bird and a penguin, but a bird that cannot fly, rather than a penguin that can?

These are the kinds of questions that a logic of reasons must address. But of course, not every question concerning reasons is a matter of logic, in this sense, and many can safely be ignored. The logic of reasons need not concern itself with the nature of reasons, for

example, just as standard logic is able to map out the logical relations among propositions without committing itself to any particular account of the nature of propositions. Nor need the subject consider the question of what reasons actually hold, or what reasons should be taken to bear on some situation, just as standard logic need not consider the question of what propositions are actually true. The logic of reasons is focused only on the question of how reasons—no matter what their nature, or where they come from—support or justify the actions or conclusions they do.

Many philosophers, particularly those working in ethics, suppose that an answer to this question is provided by what we will call the *weighing conception*, according to which reasons support actions or conclusions by contributing a kind of metaphorical weight, and that, in case of conflict, the option with the greater weight should be selected.[1] This idea can seem plausible if we limit our consideration to simple examples such as those considered so far. Why rescue Max, rather than meet Emma? Because, although you have reasons favoring both actions, your reason for rescuing Max is weightier—stronger, or more powerful—than your reason for meeting Emma. Why conclude that Tweety cannot fly, rather than that he can? Because your reason favoring the first conclusion, provided by the fact that Tweety is a penguin, carries more weight than your reason favoring the second, provided by the fact that Tweety is a bird.

The weighing conception founders, however, once we turn to situations in which the interactions among reasons are more complicated. To begin with, weighing is necessary at all only when options are incompatible—as when, for example, you believe that it is impossible both to rescue Max and meet Emma. But which options are taken to be incompatible may itself depend on background beliefs, themselves based on further reasons. You may

[1] See, for example, Baier (1958, p. 102), Harman (1975, p. 112), and Schroeder (2007, p. 130).

have arrived at your belief that rescuing Max rules out meeting Emma itself through the consideration of reasons, possibly involving the available transportation options, which had to be weighed against other reasons supporting a contrary conclusion. In that case, it would have been important for you to perform the weighing involved in deciding whether rescuing Max and meeting Emma are incompatible before deciding which of these actions to perform; you would not want to decide that you ought to perform both actions, for example, prior to reasoning about their compatibility, and then, possibly, conclude that they must be compatible because ought implies can.[2] Even in situations in which the weighing conception is plausible, then, it is clear that the process of settling on actions or conclusions through a process of weighing reasons would have to be carefully structured, with certain decisions settled before others are considered.

But there are also situations in which the interactions among reasons do not seem to involve weighing at all. Reasons support outcomes, and *defeat* other reasons, preventing them from yielding conclusions they otherwise would. There are, however, at least two kinds of defeat. The metaphor of weighing applies most naturally when, as in the situations considered so far, a reason supporting one conclusion is *rebutted* by a stronger, or weightier, reason supporting a conflicting conclusion; but, as first noted by Pollock, there are also situations in which reasons are not actually rebutted by stronger reasons supporting conflicting conclusions, but simply removed from consideration, or *undercut*.[3] To illustrate, suppose once again that Eric tells you that Tweety is a bird, providing you with a reason for concluding that Tweety is in fact a bird, but this time, what Rachel tells you is that Eric is unreliable, or at least unreliable in his ability to distinguish birds from other objects. In that case, your

[2]This process of reasoning is described as "wishful thinking" in Thomason (2000).

[3]See Pollock (1970) for an initial discussion of his distinction between rebutting and undercutting defeat; his own treatment of these ideas is developed in detail in later work, such as Pollock (1995).

reason, provided by Eric's assertion, for concluding that Tweety is a bird would be defeated, but not rebutted, since you have no stronger reason supporting the opposite conclusion—Rachael herself may insist that she does not know whether Tweety is a bird. Instead, your reason for concluding that Tweety is a bird is undercut, since it is based on Eric's assertion, but you have now learned that Eric is unreliable.

One reason might, also, simply weaken another without undercutting that reason, or removing it from consideration entirely.[4] Suppose Eric says that Tweety is a bird but Susan denies this, claiming that Tweety is not a bird—imagine that Eric and Susan are both looking at Tweety from a distance. Here, you are presented with reasons supporting conflicting conclusions, but suppose that what Rachael tells you in this case is that Eric's vision is poor. Although this does not entirely undercut your reason for concluding that Tweety is a bird, it does weaken it, perhaps to the extent that you are willing to accept the conflicting conclusion supported by Susan's statement. Or, one reason might strengthen another.[5] Suppose that, in the same situation, what Rachael now tells you is, not that Eric's vision is poor, but that Susan's is particularly good. Then, although the reason provided by Eric's statement is not weakened, the reason provided by Susan's statement is strengthened, again suggesting that you should settle on the conclusion that Tweety is a bird, though in a different way.

As these examples show, the interactions among reasons can be complex enough that it is hard to see how they could be understood through a model that allows only weighing. What other options are available, then, to help us understand the complexity? In recent years, two formal systems have been introduced with to goal of explicating the ways in which reasons interact to support the actions and conclusions they do. The first of these is the theory of defeasible reasoning developed in the seminal work of Pollock; the second is a

[4] See the discussion in Chapter 5 of Dancy (2004).

[5] Again, see Chapter 5 of Dancy (2004).

more recent theory due to Horty, which adapts and develops the default logic introduced by Reiter to provide an account of reasons.[6] Pollock's account is based on his earlier work in epistemology, and is applied, primarily, to problems in epistemology and computer science; there is a close connection between Pollock's theory of defeasible reasoning and his work on probabilistic reasoning. Horty's account begins with theories of default reasoning developed in computer science and is applied to problems in deontic logic, ethics, and legal reasoning, as well as epistemology; there is no connection to probabilistic reasoning, which seems less appropriate in the normative domain.

Setting aside incidental differences like these, however, the two accounts share some important similarities. Both are based on the idea of supplementing ordinary logic with new rules, according to which a proposition X is taken to favor, or support, an action or conclusion Y. Pollock refers both to these new rules themselves and to the premises of these rules as reasons, an ambiguity that introduces some confusion; Horty, following Reiter, describes them as default rules, or simply defaults, and refers only to their premises as reasons. There is, in both theories, an ordering on rules representing the priority, or strength, of the reasons they present, and the general idea is then to show how these rules interact with the rules of deductive inference to justify, or recommend, actions or conclusions, not just through deduction, but also on the basis of reasons. Moreover, both theories reflect the same general principle, which can be stated as follows: if the reason X supports the outcome Y, and it has been established that X, then we should accept Y as well, unless the reason X has been

[6] Pollock's initial account can be found in his (1987); this account was superseded by that of Pollock (1994), which is presented in more detail in Pollock (1995), generally taken as providing the canonical version of his theory. In fact, however, Pollock continued to refine his approach in later papers including (2001) and (2009). Horty's account is described most completely in his (2012); this account is based on the default logic introduced by Reiter (1980).

defeated—that is, either rebutted by a stronger reason supporting a conflicting conclusion, or removed from consideration as a reason, and so undercut. This idea is implemented in different ways in the two theories, though not so different that precise comparisons cannot be drawn. However, the implementations are complex enough, in both cases, to prevent useful summarization in a brief article. The presentations by the authors are accessible, in any case. And we would not want to give the impression that we think that work on the logic of reasons must follow the path mapped out in either of these theories—indeed, we feel that the field is wide open.

In the remainder of the chapter, therefore, will concentrate on a number of issues bearing on the logic of reasons that are either not treated in the work of Pollock and Horty, or whose treatment there is, we feel, either inadequate or incomplete. These are: first, the question of whether it is necessary to understand logical interactions among reasons themselves, rather than simply between reasons and the actions or conclusions they support, and if so, what principles might govern these interactions; second, priority relations among reasons and the notion of reason accrual; and third, some problems posed by undercutting defeat.

2 Relations among reasons

We have been slipping back and forth, rather casually, between what might be called *practical* and *epistemic* reasons—reasons for actions, versus reasons for conclusions. The information that Tweety is a bird might be said to provide an epistemic reason supporting the conclusion that Tweety flies. By contrast, when you promise to meet Emma for lunch, your promise is most naturally interpreted as providing a practical reason. It does not necessarily support the conclusion that you will meet Emma for lunch, but provides you with a reason for doing so.

Various theses could be advanced concerning the relation between these two kinds of reasons. One thesis is that epistemic reasons should be subsumed as a species under the genus of practical reasons. On this view, the reason for the conclusion that Tweety flies does not, in fact, support a proposition, but actually recommends an action: perhaps the action of *concluding* that Tweety flies. Another thesis is that practical reasons should be subsumed as a species under the genus of epistemic reasons. On this view, your reason to meet Emma for lunch does not recommend an action but actually supports a proposition: perhaps the proposition that you *ought* to meet Emma for lunch. Yet a third thesis is that neither practical nor epistemic reasons can be subsumed under the genus of the other, but that they are simply distinct kinds of reasons, though strikingly similar in many of their important logical properties.[7]

In addition to issues concerning relations between different kinds of reasons—practical and epistemic—there are also issues concerning logical relations among reasons. Standard logic, of course, explores logical relations among propositions, most notably the relation of entailment. The first issue, then, is whether there is any need at all to explore similar logical relations among reasons—or more starkly, are there logical relations among reasons themselves?

It is natural to assume so, and an argument can be supplied, deriving from the standard answer to the question of how reasons can support ought statements. The simplest form of this standard answer—found in Chisholm, for example—is that an agent ought to perform an action just in case there is an undefeated reason for the agent to perform that action.[8] A more general form—found in Baier, Harman, Schroeder, and many others—is that an

[7]The reader who is interested in these issues should consult [[Schroeder *this volume*]], where they are discussed in detail.

[8]See Chisholm (1964, p. 149) and (1974, p. 125).

agent ought to perform an action just in case the reasons that favor performing that action outweigh the reasons that oppose doing so.[9] Either form of the answer requires that, if an agent ought to perform an action, then that agent has a reasons to perform that action—that there is a reason corresponding to each ought.

But now, imagine a situation in which I have promised to lend $5 to a friend tomorrow and also to donate $5 to charity, so that I have reasons to do both of these things. Suppose, also, that I have only $15, with no immediate prospect of getting more, and that a movie costs $7. It seems to follow, then, that I ought not to see a movie tonight. If there is a reason corresponding to each ought, then it follows that I must have a reason not to see a movie tonight. But where did this reason come from? It is generally assumed that this new reason is derived, somehow, from my existing reasons, to give $5 each to my friend and to charity, together with certain facts about the situation in which I find myself: that I have only $15 and a movie costs $7. If one reason can, in fact, be derived from others, we can hope to discover the principles governing this entailment relation among reasons—we can hope, that is, to discover principles governing the logical relations among reasons themselves.

Against this argument, it is worth noting that neither of the two existing formal systems that can be thought of as addressing the relation between reasons and conclusions—those of Pollock and Horty—devotes any attention at all to questions concerning logical relations among the reasons themselves. Pollock does not even discuss the issue; Horty, while not denying that there may be logical relations among reasons, argues that there is no need to appeal to such relations in order to understand the way in which reasons support conclusions, either practical or epistemic.

[9]See, for example, Baier (1958, p. 102), Harman (1975, p. 112), and Schroeder (2007, p. 130).

3 Conflicts among reasons

Still, many people have found it to be either necessary or helpful to appeal to logical relations among reasons.[10] In this section, therefore, simply to illustrate the kinds of difficulties that must be faced by anyone hoping to develop a logical account of these relations, we consider one particular issue in more detail: the treatment of conflicting reasons.

We introduce this issue indirectly, by considering, first, some plausible patterns of inference in deontic logic as well as the problems they present in the presence of deontic conflicts, and then noting that both these patterns of inference and the problems they present would be relevant to a logic of reasons as well.[11]

Imagine to begin with, then, that the laws of some country set the speed limit at fifty miles per hour. From this, we can conclude that drivers ought to drive no faster than fifty miles per hour—but it seems that we can conclude also that drivers ought to drive no faster than one hundred miles per hour. This case illustrates the plausibility of the following principle of *single ought closure*, according to which we ought to perform whatever actions are entailed by the actions we ought to perform:

> If an agent ought to do X, and doing X entails doing Y, then the agent ought to do Y.

Next, suppose that the laws require, also, that all citizens must either join the military or perform alternative service, but that some particular citizen, Smith, is committed to a pacifist religion that requires her not to join the military. From this, we can conclude that

[10] A special area of interest concerns the relations among non-instrumental and instrumental reasons—among, that is, reasons deriving directly from goals, or ends, and reasons bearing on means toward those ends. This topic is discussed in greater detail by [[Kolodny *this volume*]].

[11] An encyclopaedic discussion of normative conflicts in deontic logic can be found in Goble (2014).

Smith, as a citizen, ought either to join the military or perform alternative service, but that, as a pacifist, Smith ought not to join the military. It seems to follow, then, that Smith ought to perform alternative service.

Single ought closure alone cannot explain the this desirable conclusion, since performing alternative service is not entailed by either of the single oughts established thus far: joining the military or performing alternative service, or not joining the military. The conclusion would follow, however, if we could conjoin, or agglomerate, these separate oughts together into one, since performing alternative service does follow from: joining the military or performing alternative service, and also not joining the military. What this suggests is that we need an additional principle of *consistent ought agglomeration*, which allows us to conjoin single oughts together, at least when they are consistent:

> If an agent ought to do X and the agent ought to do Y, where X and Y are consistent, then the agent ought to do X and Y.

This ought agglomeration principle and the previous ought closure principle, taken together, give us a tidy explanation of Smith's obligation. Since Smith ought to either join the military or perform alternative service, and Smith ought not to join the military, agglomeration allows us to conclude that Smith ought to either join the military or perform alternative service, and also not join the military. From this, closure then allows us to conclude that Smith ought to perform alternative service.

Now that we have seen some desirable inference principles, we are in a position to see why it is hard to reconcile the existence of conflicting obligations with these principles. Suppose you face conflicting obligations. Perhaps (1) You ought to meet Sam for a drink, but also (2) You ought not to meet Sam for a drink. The principle of single ought closure applied to (1) now yields (3) You ought to meet Sam for a drink or throw Sam into a canal. Since it is

possible to meet Sam for a drink or throw Sam into a canal and not meet Sam for a drink, we may apply consistent ought agglomeration to (2) and (3), yielding (4) You ought to meet Sam for a drink or throw Sam into a canal, and also not meet Sam for a drink. Finally, single ought closure, again, applied to (4) yields (5) You ought to throw Sam into a canal.

As this example shows, the two principles of single ought closure and consistent ought agglomeration, which are helpful for generating desirable conclusions in ordinary cases, also seem to support random, and unfortunate, conclusions when applied to conflicting oughts. And of course, one familiar reaction to this result, and others like it, is simply to deny that that conflicting obligations, or oughts, are possible—the tension between these plausible principles of inference and conflicting oughts is often taken as a reductio of the idea that there might be conflicting oughts at all. We do not intend here to try to assess this reaction to this problem of conflicting obligations is correct; instead, our goal is to show only that problem can be generalized to apply, not just to conflicting oughts, but to conflicting reasons.

The generalization is straightforward. Earlier, we imagined a country with a fifty miles per hour speed limit—from which we concluded that drivers ought to drive no faster than fifty miles per hour, and so, by ought closure, no faster than one hundred miles per hour. But it is natural to think that laws provide reasons as well as oughts, particular if we accept the idea, mentioned in the previous section, that oughts correspond to undefeated reasons, or to the most weighty sets of reasons. In this case, then, we must conclude that drivers have a reason not to drive faster than fifty miles per hour, and so a reason not to drive faster than one hundred miles per hour. But where could this latter reasons come from, since, we can assume, there is no law explicitly requiring drivers to drive less than one hundred miles per hour? By analogy to our earlier proposal, it is plausible to suppose that this second reason is derived from the first through a principle of *single reason closure*, according to which:

> If an agent has a reason to do X, and X entails Y, then the agent has a reason to do Y.

Let us return also to Smith, subject to the laws of her country, but also to pacifist commitments. Once again, it is natural to conclude that she has a reason, based on the laws, either to join the military or perform alternative service and another reason, based on her pacifism, not to join the military—and so, it seems, a reason to perform alternative service. But again, this latter reason does not follow from the others unless we can assume, in addition, a principle of *consistent reason agglomeration*, according to which:

> If an agent has a reason to do X and the agent has a reason to do Y, where X and Y are consistent, then the agent has a reason to do X and Y.

With these two principles, just as before, we can understand how Smith's latter reason can be derived from the others. Reason agglomeration allows us to conclude, from Smith's separate reasons, first, to join the military or perform alternative service, and second, not to join the military, that Smith has, in addition, a single reason either to join the military or perform alternative service, and also not to join the military. And then from this, single reason closure yields the conclusion that Smith has a reason to perform alternative service.

But since these two principles about reasons are structurally similar to the earlier principles about oughts, and the earlier principles yielded peculiar results when applied to conflicting oughts, it is easy to see that the current principles generate equally peculiar results when applied to conflicting reasons. Suppose, for example, that you have promised Sam to meet him for a drink, but also that you have promised Melissa not to meet Sam for a drink. Then you have reasons to meet Sam for a drink, and also not to meet Sam for a drink—and it is now possible, based on our principles, to mimic for reasons exactly the same argument

set out earlier for oughts to reach the conclusion that you have a reason to throw Sam into a canal.

Now, as we mentioned, one popular reaction to this kind of problem in the case of oughts is to conclude simply that there can be no conflicting oughts—since the possibility of conflicting oughts, taken together with our very plausible principles, would yield such strange results. However, the analogous reaction is not available in the case of reasons: the whole point of reasons is that they can support an outcome without guaranteeing that outcome, thus allowing for different reasons to support conflicting outcomes, and to do so consistently.

Because the option of simply denying the possibility of conflicting reasons is not available, the tension with our reasoning principles is much more serious here than in the case of conflicting oughts. Here, with reasons, there is no choice other than to reject either reason closure or reason agglomeration. And this may be easy enough to do. What is harder, though, is to reject these principles while still allowing the patterns of inference they were originally introduced to support. Those writers, such as Pollock or Horty, who are able to proceed without positing logical relations among reasons can avoid this problem entirely; but those whose practical aims or theoretical commitments require logical relations among reasons must, eventually, confront the task of mapping out those relations.[12]

[12] As we have emphasized, the problem of conflicting reasons is structurally analogous to the problem of conflicting obligations in deontic logic. This suggests that it is promising to think that some of the proposals recently developed for handling conflicts within deontic logic might be adapted to handle conflicts among reasons as well. See, again, Goble (2014) for a survey of recent treatments of conflicting oughts within deontic logic; see [[Nair ??]] for an attempt to adapt one of these to the problems presented by conflicting reasons.

4 Weights and priorities

As we have seen, reasons are thought to come with weights, or more generally, to be arranged in a priority ordering. Two questions now arise concerning the priority relations among reasons: first, where do these the priorities come from, and second, what properties can the overall priority ordering be expected to satisfy?

Priorities among reasons can different sources. In our initial penguin example, the priority between the two reasons provided by the fact that Tweety is both a bird and a penguin—supporting the conflicting conclusions that he can fly, and that he cannot—seems to derive from specificity: a penguin is a specific kind of bird, and so information about penguins in particular takes precedence over information about birds in general. But even in the epistemic domain, there are priority relations that have nothing to do with specificity. Reliability is another source. Suppose that Susan, a meteorologist, tells you that it will rain today, but that Eric, a mere person in the street, tells you that it will not; here, although both assertions provide reasons, it would be natural for you to view Susan as more reliable on this topic, and so to assign greater weight, or higher priority, to the reason provided by her assertion. And once we move from epistemic to practical reasons, then authority provides yet another source for priority relations. National laws typically override state or provincial laws, and more recent court decisions have more authority than older decisions; direct orders override standing orders, and because a Colonel outranks a Major, orders from the Colonel override orders from the Major.

Finally, one of the most important sources of priority is our very own reasoning about which reasons should have higher priority than others Just as we reason about ordinary things in the world—birds, penguins, the weather—so we reason about our own reasons, offering further reasons for taking some of our reasons more seriously than others, and still

further reasons for evaluating those. This is particularly evident in well-structured normative domains, such as the law, where the resolution of a dispute often involves explicit arguments concerning the relative importance of different considerations bearing on some issue. But it occurs also in the epistemic domain. Consider reliability, again. Suppose, in our previous example, that someone disagreed with your view that Susan is more reliable than Eric concerning the weather. What would you say to such a person? Most likely, you would offer further reasons for favoring the reliability of the meteorologist, and these reasons might themselves have to be buttressed by appeal to further reasons still.

A logic of reasons should provide an account of the way in which actions and conclusions are supported by reasons together with their priorities, while at the same time, the priority relations among reasons are established by the same process of reasoning they serve to guide. This may sound complicated—perhaps forbiddingly so, perhaps circular—but it turns out to be doable, and perhaps in more than one way. The issue was not addressed by Pollock, who adopts the common view that the weight of a reason is fixed, either by convention or as part of its nature, not something that is itself to be established through reasoning.[13] But a manageable account is provided by Horty, based on previous work in the field of artificial intelligence and law.[14] A different approach is sketched by Schroeder.[15]

We turn now to the properties that a priority ordering on reasons, whatever its source, might be expected to satisfy. Where r, r', and r'' are reasons, it is natural, first of all, to

[13]See, for example, Baier (1958, p. 106) for the suggestion that the strength of reasons is determined by the "social environment" as well as Skorupski (1997), who works with a model of reasons according to which the weight of a reason is, in a sense, part of its nature.

[14]See Chapter 5 of Horty (2012). The particular approach presented there draws on techniques introduced by Gordon (1993) in his analysis of legal reasoning; these techniques have been refined and developed by a number of people, notably including Brewka (1994) as well as Prakken and Sartor (1995; 1996).

[15]See the treatment of "weight recursion" in Chapter 5 of Schroeder (2007).

suppose that this ordering should be *transitive*, so that, if r' has a higher priority than r and r'' has a higher priority than r', then r'' has a higher priority than r; and it seems equally natural to assume that the ordering satisfies the property of *irreflexivity*, according to which it is not the case that r has a higher priority than r. An ordering relation that is both transitive and irreflexive is referred to as a *strict partial ordering*.

Should we assume any other properties in the priority orderings? In particular, should we assume that this ordering satisfies the property of *connectivity*, according to which we could suppose, for any distinct reasons r and r', that either r or r' has a higher priority than the other? Here we reach an important branch point in our discussion. On one hand, this connectivity assumption would allow for a straightforward resolution to any potential conflicts among default rules. What the assumption tells us is that, of any two such rules, and their corresponding reasons, one is always stronger than the other; and so it would be natural, in case of a conflict, to settle the matter simply by favoring the stronger of the two. On the other hand, connectivity is not particularly plausible, in either the practical or the epistemic domain, and for two reasons.

First of all, some reasons seem simply to be incommensurable. The canonical example in the practical domain is Jean-Paul Sartre's description of a student during the Second World War who felt for reasons of patriotism and vengeance (his brother had been killed by the Germans) that he ought to leave home to fight with the Free French, but who also felt, for reasons of sympathy and personal devotion, that he ought to stay at home to care for his mother.[16] Sartre presents the situation in a vivid way that really does make it seem as if the reasons confronting the student derive from entirely separate sources of value—duty to country, versus duty to family—and cannot meaningfully be compared in importance.

[16]The example is from Sartre (1946).

It is more difficult to construct a plausible example of incommensurability in the epistemic domain, but consider a hypothetical election between two candidates in an isolated southern congressional district. Suppose that the Associated Press poll strongly favors Candidate 1 while an experienced local politician, who is neutral in the contest, confidently predicts a victory for Candidate 2. Modern statistical polling is exacting and scientific, but it can yield incorrect results in unfamiliar situations, where parameters might not have been set properly to reflect local circumstances. The politician, by contrast, has an intuitive and historically-grounded sense of his community, which we can suppose he has relied on throughout his successful career, but opinions might have shifted in ways of which he is not aware, and which a survey would detect. A situation like this, then, would seem to present us with conflicting epistemic reasons, deriving from entirely different sources, which are at least arguably incomparable.

The second reason for questioning connectivity is that conflicting reasons are in fact comparable in strength, they might nevertheless violate the connectivity property, according to which one or the other must be stronger, simply by having equal strength. As an example from the practical domain, suppose you have inadvertently promised to have a private dinner tonight with each of two identical and identically situated twins, both of whom would now be equally disappointed by my cancellation; the situation can be made arbitrarily symmetrical.[17] You now have two reasons, based on your promises, that favor two conflicting actions: having a private dinner with Twin 1, or having a private dinner with Twin 2. Since these reasons are of exactly the same kind, they can meaningfully be compared in priority. But given the symmetry of the situation, it is hard to see how either should be assigned a higher priority than the other.

[17]The importance of symmetrical cases like this in practical reasoning was emphasized by Marcus (1980).

It is, again, a bit harder to find convincing examples of equally weighted, conflicting reasons in the epistemic domain, but consider this. Suppose you plan to have lunch with two friends, at a place selected by the two of them. Imagine that, as lunch time approaches, you receive two text messages, one from each friend. The first reads, "We'll be at Mark's Kitchen, meet us there," and the second, "We'll be at Everyday Gourmet, meet us there." A check of the time stamps shows that the two messages were sent simultaneously—and at this point your phone battery dies, so that you can neither inquire nor receive any further corrections. You are now faced with a practical decision: Where should you go? But the answer hinges on an epistemic question: Where will your friends be? And in considering this questions you seem to be faced once again with two reasons, the text messages sent by your equally reliable friends, identical in strength but supporting conflicting conclusions.

5 Reason accrual

In light of the apparent counterexamples to connectivity, it seems tempting to conclude that the priority relation on reasons satisfies only the two strict partial ordering constraints, transitivity and irreflexivity, allowing for the possibility of reasons that are either incomparable or identical in priority. This conclusion would challenge the metaphor according to which the strength or force of a reason can be taken as a kind of weight, since it is so central to the notion of a weight that two weights can be compared, or combined. There are also, however, arguments that support the metaphor of reasons as possessing weights. The most powerful of these centers around the idea of reason accrual.

Consider an example: Max been invited to the wedding of a distant relative at a difficult time of year. He is not particularly close to this relative, and, since the wedding falls at such an inconvenient time, would rather not go. But suppose he learn that the guests will

include his two old aunts, Olive and Petunia, whom he enjoys and who he knows would like to see him. Here it is perfectly sensible to imagine that, even though Max would choose not to attend the wedding if only one of the two aunts were going, the chance to see both Olive and Petunia in the same trip offers enough value to compensate for the inconvenience of the trip itself.

How can this example be understood from the standpoint of reasons? It is simplest to assume that Max is working with three reasons: the inconvenience of the trip, which favors not going, the prospect of seeing Olive, which favors going, and the prospect of seeing Petunia, which favors going. If we suppose that the inconvenience of the trip has higher priority than the prospect of seeing either aunt individually , then it appears that he should not go, since each of the reasons he has for going—seeing Olive, seeing Petunia—is defeated by a stronger reason not to. But of course, this is the wrong result. Although the inconvenience of the trip would outweigh the benefit of seeing either Olive or Petunia individually, it does not outweigh the benefit of seeing them together.

This is an example which the model that assigns weights to reasons fits very well. One can almost picture a scale on which the inconvenience of the trip is balanced against, and outweighs, the benefits of seeing Aunt Olive alone, and then watching as the value of seeing Aunt Petunia as well is added to the scale and the balance slowly tips to the other side, favoring the trip after all. The problem with the picture that assigns to reasons only priorities is that it does not explain how the priorities associated with a number of different reasons can be combined—much as weights can be combined—to defeat another reason that is capable of defeating each of the initial reasons, considered individually.

We might, of course, assume that the example should be understood as presenting Max with a fourth reason: not just the prospect of seeing Olive and the prospect of seeing Petunia,

but the prospect of seeing both Olive and Petunia—which also favors going to the wedding, and which, we can suppose, carries a higher priority than the inconvenience of the trip. We would then reach the correct conclusion that Max should go to the wedding, since his only reason for not going, the inconvenience of the trip, would be defeated by a stronger reason for going.

In a way, though, this representation of the example simply relocates the issue. The question, now, concerns the relation between this new reason for going to the wedding, the presence of Olive and Petunia, and the previous reasons provided by the presence of Olive, and the presence of Petunia. One option is to suppose that the new reason is entirely independent of the previous two. This option has the advantage of simplicity, and, as its sole disadvantage, a significant degree of implausibility: can we really say that the reason arising from the presence of Olive and Petunia together has nothing at all to do with the separate reasons arising from the presence of Olive, and of Petunia?

The other option, of course, is to suppose that the complex reason arising from the presence of Olive and Petunia together, along with its place in the priority ordering, is derived, somehow, from the simpler reasons provided the presence of Olive and of Petunia, along with their priorities. This option carries more initial plausibility, but raises a number of difficult issues. Surely, not just any two reasons favoring some outcome can be combined to form a stronger reason favoring that outcome, and sometimes reasons that favor a particular outcome when taken individually favor another outcome entirely when taken together. Suppose, for example, that Symptom 1 is a reason for the administration of Drug A, since it suggests Disease 1, for which Drug A is appropriate, and that Symptom 2 is also a reason for the administration of Drug A, since it suggests Disease 2, for which Drug A is also appropriate; still, it might be that Symptoms 1 and 2 appearing together suggest Disease 3, for which

Drug A is not appropriate.[18]

If reasons can indeed be composed into complexes, it is clear that the behavior of the complexes is not easily predictable from that of the components: some complex reasons, such as seeing Olive and Petunia together, form stronger reasons supporting the same conclusions as their components, while others, such as Symptoms 1 and 2, actually interfere with the support provided by their component reasons. Anyone favoring the option of reason composition would therefore, at some point, have to provide a careful description of the process, and of the constraints governing this process. Which reasons can properly be composed and which cannot? And how is the priority of a complex reason influenced by the priority of its components? These are difficult questions, and we are not aware of any successful attempt to answer them in the literature.[19]

6 Undercutting defeat

A reason is undercut when it is, simply, removed from consideration—defeated, without being rebutted by any stronger reason supporting a conflicting conclusion. The concept was illustrated earlier by the situation in which Eric asserts that Tweety is a bird, but Rachel asserts that Eric is unreliable. Here, Rachel's assertion can be seen as undercutting Eric's assertion as a reason for concluding that Tweety is a bird, removing it from consideration,

[18] As Dancy (2004, p. 15) writes, "reasons are like rats, at least to the extent that two rats that are supposedly on the same side may in fact turn and fight among themselves."

[19] Among writers who have discussed this issue, most take the first of our two options: Pollock (1995, pp. 101–102) treats complex reasons, along with their priorities, as independent of the simpler reasons that might naturally be taken as their components; Schroeder (2007, Chapter 7) works with a model in which weights are assigned directly to sets of reasons, rather than to individual reasons, but again does not relate the weight of a set of reasons to the weight of the singleton sets constructed from the reasons belonging to that set. [[Cite Nair article in weighing reasons book.]]

and so defeating it, without providing any reason at all for the contrary conclusion that Tweety is not a bird.

The relations between undercutting defeat and ordinary rebutting defeat can be complicated. In our previous example, Rachael's assertion functioned as an undercutter, providing a reason for removing Erics's assertion from consideration as a reason supporting the conclusion that Tweety is a bird. But suppose that Susan now asserts that Rachael is unreliable. In that case, Susan's assertion would function as an undercutter undercutter, providing a reason for removing Rachael's assertion from consideration as a reason for removing Eric's assertion from consideration as a reason for the conclusion that Tweety is a bird, and so allowing Eric's assertion to re-emerge as a reason. By contrast, suppose that Susan, who is more reliable than Rachael, had contradicted Rachael's statement by asserting that Eric is, in fact, reliable. Susan's assertion would then function as an undercutter rebutter, defeating Rachael's assertion by providing a stronger reason supporting a conflicting conclusion. Having been provided with simple examples of undercutter undercutters and undercutter rebutters, the reader is now invited to fashion for him or herself examples of rebutter undercutters, rebutter rebutters, and even more elaborate examples.

Although the concept of undercutting defeat was first highlighted by Pollock, as mentioned earlier, the idea, or a very similar idea, is also discussed in the literature on practical reasoning, where it is considered as part of the general topic of *exclusionary* reasons, introduced by Raz.[20] Raz motivates the concept through a number of examples, but we consider here only the representative case of Colin, who must decide whether to send his son to a

[20]See Section 1.2 of Raz (1975) for his initial discussion. The topic has spawned an extensive secondary literature, notably including papers by Gans (1986), Moore (1989), and Perry (1987; 1989), subsequent elaborations by Raz himself, both in his (1989) and in a postscript to the second edition of his (1975), and a review of the latter by Edmundson (1993).

private school. We are to imagine that there are various reasons pro and con. On one hand, the school will provide an excellent education for Colin's son, as well as an opportunity to meet a more varied group of friends; on the other hand, the tuition is high, and Colin is concerned that a decision to send his own son to a private school might serve to undermine his support for public education more generally.

However, Raz asks us to imagine also that, in addition to these ordinary reasons pro and con, Colin has promised his wife that, in all decisions regarding the education of his son, he will consider only those reasons that bear directly on his son's interests. And this promise, Raz believes, cannot properly be viewed as just another one of the ordinary reasons for sending his son to the private school, like the fact that the school provides a good education. It must be viewed, instead, as a reason of an entirely different sort—a "second-order" reason, according to Raz, for excluding from consideration all those ordinary, or "first-order," reasons that do not bear on the interests of Colin's son. Just as, once Rachel tells you that Eric is unreliable, you should disregard Eric's statement as a reason for concluding that Tweety is a penguin, Colin's promise should lead him, likewise, to disregard those reasons that do not bear on the interests of his son. An exclusionary reason, on this interpretation, is nothing but an undercutting defeater in the practical domain.

Given a collection of reasons, some of which undercut, or exclude, others, how are we to determine what actions or conclusions are supported? In fact, the logic mentioned earlier, due to Pollock and by Horty, both address this issue. Pollock's theory has broader coverage, since it provides an account of certain collections of reasons that Horty's theory fails to treat. For example, suppose Eric asserts that he is unreliable, so that his assertion provides us with a reason for the conclusion that he is unreliable, which is itself, then, an undercutter—a reason for removing his assertion from consideration as a reason for the conclusion that he is

unreliable. Pollock's theory provides an account of this example; Horty's does not. On the other hand, Horty's theory offers a unified treatment of undercutting defeat in epistemology and exclusion in practical reasoning. In addition, since Horty allows reasoning about the priority of reasons, he is able to explore certain issues about how reasoning about priority interacts with reasoning about exclusion.

Pollock and Horty also differ about one narrow, but very interesting issue.[21] It is part of the idea of ordinary rebutting defeat, of course, that a reason supporting a particular action or conclusion cannot be rebutted by a weaker reason. For example, if Eric says that Tweety is a bird and Rachael says that he is not, then if Rachael is less reliable than Eric, the reason supplied by Rachael's statement does not rebut that supplied Eric's. But can a reason be undercut, or excluded, by a weaker reason? Where Rachael remains less reliable than Eric, again suppose that Eric says that Tweety is a bird, but that what Rachael says, this time, is that Eric is unreliable. Can the reason supplied by Rachael's statement, in this case, undercut the reason supplied by Eric's, even though it is weaker? In a way, it is hard to see why strength of reasons should matter at all, since Eric and Rachael are not even talking about the same things—Eric is talking about whether or not Tweety is a bird, while Rachael is talking about the entirely different topic of whether or not Eric is trustworthy. Nevertheless, it can seem odd that the less reliable Rachael should be able to undercut the more reliable Eric. Pollock offers an argument for why undercutters can be no weaker than the reasons they undercut, and builds this condition into his definition of undercutting defeat. Horty abandons this condition, questions Pollock's argument, and attempts to explain our intuitions of oddity in a different way. The issue is, again, narrow, and somewhat technical, but it gives an indication of the kind of detailed question that anyone hoping to develop a

[21] See Pollock (1995, pp. 103–104) and Horty (2012, pp. 135–141).

logic of reasons must address.

References

Baier, Kurt (1958). *The Moral Point of View: A Rational Basis of Ethics*. Cornell University Press.

Brewka, Gerhard (1994). Reasoning about priorities in default logic. In *Proceedings of the Twelveth National Conference on Artificial Intelligence (AAAI-94)*, pages 940–945. AAAI/MIT Press.

Chisholm, Roderick (1964). The ethics of requirement. *American Philosophical Quarterly*, 1:147–153.

Chisholm, Roderick (1974). Practical reason and the logic of requirement. In Körner, Stephan, editor, *Practical Reason*, pages 2–13. Blackwell Publishing Company. Reprinted in Raz (1978).

Dancy, Jonathan (2004). *Ethics Without Principles*. Oxford University Press.

Edmundson, William (1993). Rethinking exclusionary reasons: a second edition of Joseph Raz's 'Practical Reason and Norms'. *Law and Philosophy*, 12:329–343.

Gans, Chaim (1986). Mandatory rules and exclusionary reasons. *Philosophia*, 615:373–394.

Goble, Lou (2014). Prima facie norms, normative conflictrs, and dilemmas. In Gabbay, Dov, Horty, John, Parent, Xavier, van der Meyden, Ron, and van der Torre, Leendert, editors, *Handbook of Deontic Logic and Normative Systems*, pages 241–351. College Publications.

Gordon, Thomas (1993). *The Pleadings Game: An Artificial-Intelligence Model of Procedural Justice*. PhD thesis, Technische Hochschule Darmstadt.

Harman, Gilbert (1975). Reasons. *Critica*, 7:3–18. Reprinted in Raz (1978); pagination refers to this version.

Horty, John (2012). *Reasons as Defaults*. Oxford University Press.

Marcus, Ruth Barcan (1980). Moral dilemmas and consistency. *Journal of Philosophy*, 77:121–136.

Moore, Michael (1989). Authority, law, and Razian reasons. *Southern California Law Review*, 62:827–896.

Perry, Stephen (1987). Judicial obligation, precedent, and the common law. *Oxford Journal of Legal Studies*, 7:215–257.

Perry, Stephen (1989). Second-order reasons, uncertainty, and legal theory. *Southern California Law Review*, 62:913–994.

Pollock, John (1970). The structure of epistemic justification. In *Studies in the Theory of Knowledge*, American Philosophical Quarterly Monograph Series, number 4, pages 62–78. Basil Blackwell Publisher, Inc.

Pollock, John (1987). Defeasible reasoning. *Cognitive Science*, 11:481–518.

Pollock, John (1994). Justification and defeat. *Artificial Intelligence*, 67:377–407.

Pollock, John (1995). *Cognitive Carpentry: A Blueprint for How to Build a Person*. The MIT Press.

Pollock, John (2001). Defeasible reasoning with variable degrees of justification. *Artificial Intelligence*, 133:233–282.

Pollock, John (2009). A recursive semantics for defeasible reasoning. Unpublished manuscript.

Prakken, Henry and Sartor, Giovanni (1995). On the relation between legal language and legal argument: assumptions, applicability, and dynamic priorities. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law (ICAIL-95)*. The ACM Press.

Prakken, Henry and Sartor, Giovanni (1996). A dialectical model of assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law*, 4:331–368.

Raz, Joseph (1975). *Practical Reasoning and Norms*. Hutchinson and Company. Second edition with new Postscript printed in 1990 by Princeton University Press, and reprinted by Oxford University Press in 2002; pagination refers to the Oxford edition.

Raz, Joseph, editor (1978). *Practical Reasoning*. Oxford University Press.

Raz, Joseph (1989). Facing up: a reply. *Southern California Law Review*, 62:1153–1235.

Reiter, Raymond (1980). A logic for default reasoning. *Artificial Intelligence*, 13:81–132.

Sartre, Jean-Paul (1946). *L'Existentialisme est un Humanisme*. Nagel. Translated as "Existentialism is a Humanism" in W. Kaufmann, editor, *Existentialism from Dostoevsky to Sartre*. Meridian Press, 1975.

Schroeder, Mark (2007). *Slaves of the Passions*. Oxford University Press.

Skourpski, John (1997). Reasons and reason. In Cullity, Garret and Gaut, Berys, editors, *Ethics and Practical Reason*, pages 345–368. Oxford University Press.

Thomason, Richmond (2000). Desires and defaults: a framework for planning with inferred goals. In *Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR-2000)*, pages 702–713. Morgan Kaufmann Publishers.