The Philosophical Quarterly Vol. 73, No. 1 *2023*
ISSN 0031-8094 https://doi.org/10.1093/pq/pqac024
Advance Access Publication 4th July 2022

OUTLINE OF A THEORY OF REASONS

BY VINCENZO CRUPI AND ANDREA IACONA

This paper investigates the logic of reasons. Its aim is to provide an analysis of the sentences of the form 'p is a reason for q' that yields a coherent account of their logical properties. The idea that we will develop is that 'p is a reason for q' is acceptable just in case a suitably defined relation of incompatibility obtains between p and $\neg q$. As we will suggest, a theory of reasons based on this idea can solve three challenging puzzles that concern, respectively, contraposing reasons, conflicting reasons, and supererogatory reasons, and opens a new perspective on some classical issues concerning non-deductive inferences.

Keywords: reason, conditional, evidence, support, supererogatory, contraposition.

I. PRELIMINARY CLARIFICATIONS

Our investigation concerns epistemic reasons, that is, reasons for belief. For any two propositions p and q, to say that p is a reason for q in the sense that matters to us is to say that assuming p provides a justification for believing q. In other words, p is a reason for q when p supports q. Epistemic reasons are often treated as distinct from practical reasons, that is, reasons for action, and there are different views on the relationship between these two kinds of reasons. Here we will simply restrict consideration to epistemic reasons without addressing the question whether the claims that we make about them apply to practical reasons as well.

We will use the triangle $\rhd$ to represent the relation of support that obtains between p and q when p is a reason for q. As it emerges from the explanation provided above, we take p and q to be propositions, which is quite convenient for the purposes of formal semantics. But other ways to understand the terms of the relation of support are compatible with our notation. Sometimes, 'reason' is used to denote a true proposition rather than a mere proposition, that is, a state of affairs that actually obtains. If one adopts this reading, which is stronger than ours, then one can construe $p \rhd q$ as 'p would be a reason for q', and express the claim that p is a reason for q by the conjunction of $p \rhd q$ and p.[1]

The understanding of support that characterizes the notion of reason is typically conveyed by verbs such as 'indicates', 'suggests', or 'implies', as in the following examples:

(1) White smoke indicates that a new pope has been chosen.
(2) Sophie's being French suggests that she can read French.
(3) Sophie's capacity to read French implies that she can read.

(1)–(3) are plausibly described as sentences of the form $p \rhd q$, given that their initial noun phrases express conditions that can be stated propositionally: white smoke raises, Sophie is French, Sophie can read French.

Our use of $\rhd$ is constrained by three basic assumptions, which may be regarded as minimal adequacy conditions for a theory of reasons. The first assumption is that statements about reasons can be expressed in conditional form. Here are some obvious conditional counterparts of (1)–(3):

(4) If white smoke raises, a new pope has been chosen.
(5) If Sophie is French, she can read French.
(6) If Sophie can read French, she can read.

As is well known, conditionals can be understood in more than one way, and it is an open question whether there is a unique correct theory of them. But it is hardly disputable that, at least on some intelligible reading of 'if', conditionals can be used to make statements about reasons, as in the case of (4)–(6).

The second assumption is that there is an essential conceptual link between reasons and arguments. To say that p is a reason for q is to say that the inference from p to q is justified. As long as the term 'valid' is used in a sufficiently broad way, which is not restricted to deductive validity, this amounts to saying that p is a reason for q when the argument formed by p as premise and q as conclusion is valid. For example, the following are argument counterparts of (1)–(3):

(7) White smoke raises; therefore, a new pope has been chosen.
(8) Sophie is French, so she can read French.
(9) Sophie can read French; it follows that she can read.

As long as 'p is a reason for q' is understood as 'the inference from p to q is justified', it is plausible to expect that our claims about reasons agree with the corresponding claims about inferences. In particular, we will assume that p is a reason for q whenever q logically follows from p. Consider, for example, the claim that p is a reason for p, that is, $p \rhd p$. Although this claim is not interesting

[1] Goble (2013: 266), suggests this interpretation for a similar symbol.

or informative, we would not say that it is false. Saying so would amount to denying that one can infer p from p, in spite of the fact that p logically follows from p.[2] More generally, our second assumption entails that some claims about reasons are trivial or uninformative, just like the corresponding claims about inferences. In this respect, we depart from attempts to characterize the meaning of the word 'reason' as ruling out such limiting cases.

The third assumption is that reasons can be fallible, in the sense that being a reason for q by itself does not guarantee that q is true. For example, the reason stated in (1) is fallible: an accidental combustion in the chimney of the Sistine Chapel could produce white smoke while the cardinals are still discussing. The same goes for (2): Sophie could be illiterate because she never went to school. One common way to spell out the fallibility intuition is to say that reasons can be *defeasible*, in the sense that they do not conform to the principle of *Monotonicity*:

M $p \rhd q$ entails $(p \wedge r) \rhd q$.

Non-defeasible—or conclusive—reasons may be treated as a special kind of reasons that obey Monotonicity. For example, the reason stated in (3) is conclusive. To say that p non-defeasibly supports q is to say that there is no r such that $p \rhd q$ but not $(p \wedge r) \rhd q$. By contrast, defeasibility amounts to the existence of such an r, which is called a *defeater*. In the case of (1), the relevant r can be 'There is an accidental combustion in the chimney of the Sistine Chapel,' while in the case of (2) it can be 'Sophie never went to school'.[3]

Another familiar way to spell out the fallibility intuition is to say that reasons are *non-factive*: one may have a reason for q even though q is actually false. Sometimes one may simply be unlucky. In epistemology, it is quite common to distinguish knowledge from mere justification, assuming that the latter, unlike the former, lacks factivity: While knowing that q entails that q is true, having a justification for q does not entail that q is true. Obviously, this second way to construe fallibility is directly related to the first, because it is only in connection with defeasible reasons that unlucky cases of the sort considered can arise. Conclusive reasons rule out falsity by definition.

The question that will be addressed in the following sections is how $\rhd$ is to be defined in order to provide a coherent account of the logical properties of the sentences of the form $p \rhd q$. As far as we can see, this question has not

[2] This issue is addressed, for example, in Fuhrmann (2017).

[3] A further distinction that is sometimes drawn, which goes back to Pollock (1970: 73–74), is between *rebutting* defeaters and *undercutting* defeaters. A defeater r for $p \rhd q$ is rebutting if it constitutes a reason against q, that is, a reason for $\neg q$, while it is undercutting if it questions the connection between p and q. For example, 'Sophie never went to school' supports 'Sophie cannot read French', so it is a rebutting defeater for (2). Instead, 'There has been an accidental combustion in the chimney of the Sistine Chapel' questions the connection between 'White smoke rises' and 'A new pope has been chosen', so it is an undercutting defeater for (1).

yet received the attention it deserves. At least three formal theories of reason have been developed, which converge on the three basic assumptions outlined above. One is the theory of defeasible reasoning offered by Pollock, which is based on his work in epistemology.[4] The other is the account of default reasoning due to Reiter and Horty, which relies on default logic as employed in computer science.[5] The third is the ranking-theoretic explication of reasons due to Spohn.[6] Although the three theories just mentioned provide important insights into the logic of reasons, there is still much to be said on this topic, or so we believe.

II. INTERPRETATIONS OF THE TRIANGLE

Since statements about reasons are commonly expressed by means of conditionals, as in (4)–(6), it is natural to wonder whether some extant theory of conditionals is able to provide the desired interpretation of $\triangleright$. This section presents four candidates that might be regarded as initially plausibile and sketches our own interpretation as a fifth option.

Although there is a wide variety of theories of conditionals on the market, only some of them can suit our purposes. In particular, two traditional and widely debated views must be ruled out from the very beginning. One is the material conditional view, according to which a conditional is true when it is not the case that its antecedent is true and its consequent is false. The other is the strict conditional view, according to which a conditional is true when it cannot be the case that its antecedent is true and its consequent is false. Since both views entail that conditionals are monotonic, neither of them can work as an interpretation of $\triangleright$, given what has been said above about the defeasibility of reasons. If one wants to find a theory of conditionals that provides a suitable interpretation of $\triangleright$, then one must look somewhere else.

At least four well-known theories of conditionals are compatible with the assumptions stated in the previous section. The first is the probabilistic view developed by Adams, which defines the acceptability of a conditional as a function of the conditional probability of its consequent given its antecedent. To adopt this view as an interpretation of $\triangleright$—call it *conditional probability interpretation*—is to say that $p \triangleright q$ is acceptable to the extent that $P(q|p)$ is high.[7]

The second option is the possible-world view advocated by Stalnaker and Lewis. On this view, a conditional is true when its consequent holds in the closest world, or worlds, in which its antecedent holds. To adopt this view as

[4] Pollock (1987) and Pollock (1995).
[5] Reiter (1980), Horty (2007), and Horty (2012).
[6] Spohn (2012, ch. 6).
[7] Adams (1965).

an interpretation of $\triangleright$—call it *Stalnaker-Lewis interpretation*—is to say that $p \triangleright q$ is acceptable if and only if q holds in the closest world, or worlds, in which p holds.[8]

The third option is the belief revision view elaborated by Gärdenfors and others. This view defines conditionals as acceptable relative to belief states, understood as deductively closed sets of sentences. In this interpetation—call it *belief revision interpretation*—$p \triangleright q$ is acceptable relative to a belief state K if and only if $q \in f(K, p)$, where f is a function that takes belief states and sentences as arguments and yields revised belief states as values.[9]

The fourth option is the theory of 'difference-making' conditionals suggested by Rott and embedded in Spohn's theory of reasons. In this interpretation—call it *difference-making interpretation*—$p \triangleright q$ is acceptable if and only if the following conditions are satisfied: (i) q holds in the closest worlds in which p holds, and (ii) it is not the case that q holds in the closest worlds in which $\neg p$ holds. While (i) expresses the Ramsey Test, which underlies the first three interpretations, (ii) is a further condition intended to capture the intuition that q holds in virtue of p.[10]

Note that one basic point on which these four interpretations converge is that they assume that the relation expressed by $\triangleright$ is insensitive to purely hyperintensional variations in the relata. Therefore, they all validate the classical rule of *Substitution of Equivalents*. This assumption will be retained throughout the paper.[11]

In the next sections, we will contrast the four options just presented with a fifth option, our favourite interpretation. We call it *evidential interpretation* because it is based on the evidential account of conditionals advocated by Crupi and Iacona. The core idea is that $p \triangleright q$ is acceptable if and only if p and $\neg q$ are incompatible. This incompatibility condition can be spelled out both in modal terms and in probabilistic terms. The modal version of the account—which is directly comparable with the last three options considered above—implies the following: (i) $\neg q$ does not hold in the closest worlds in which p holds and (ii) p does not hold in the closest worlds in which $\neg q$ holds. As in the case of the difference-making interpretation, (i) expresses the Ramsey Test, while (ii) is a further condition intended to capture the intuition that q holds in virtue of p. The probabilistic version of the account defines the degree of incompatibility between p and $\neg q$ in terms of the following measure, provided that $P(p \wedge \neg q)$

[8] Stalnaker (1991) and Lewis (1973).

[9] Gärdenfors (1988), Levi (1988), and Arlo-Costa (1995).

[10] Rott (1986), Rott (2022), and Spohn (2015). Rott and Spohn rely on the AGM formalism and ranking functions, respectively, as their favourite technical machinery. However, recasting their view in a possible world semantics is immaterial for our purposes, and favours uniformity and comparability with the other options under scrutiny. See Raidl (2021) for a discussion.

[11] See Faroldi & Protopopescu (2019) for a hyperintensional approach to some logical properties of reasons.

$\leq P(p)P(\neg q)$:

$$1 - \frac{P(p \wedge \neg q)}{P(p)P(\neg q)}.$$

Otherwise, the degree of incompatibility between p and $\neg q$ is 0. For the limiting cases in which $P(p) = 0$ or $P(q) = 1$, incompatibility is assumed to be maximal. The acceptability of $p \rhd q$ can then be equated with the degree of incompatibility between p and $\neg q$.[12]

III. CONTRAPOSITION

In order to compare the interpretations of $\rhd$ outlined in the previous section, we will discuss three puzzles that pose interesting challenges to any theory of reasons. The first puzzle concerns *Contraposition*, the principle stated as follows:

C $p \rhd q$ entails $\neg q \rhd \neg p$.

As far as our initial examples are concerned, Contraposition seems fine. Consider (1). If the presence of white smoke constitutes a reason for thinking that a new pope has been chosen, then it is plausible that the lack of a decision in the Sistine Chapel constitutes a reason for thinking that no white smoke arises. The following conditional seems as compelling as (4):

(10) If the new pope has not been chosen, then there is no white smoke.

Similar remarks hold for (2) and (3). From 'Sophie cannot read French' one can infer 'Sophie is not French', and from 'Sophie cannot read' one can infer 'Sophie cannot read French'.[13]

A first consideration in support of the Contraposition goes as follows. It is plausible to assume that, when one has a reason for q, one thereby has a reason against $\neg q$. That is, 'p is a reason for q' seems to entail 'p is a reason against $\neg q$'. Moreover, it is equally plausible to assume that, if one has a reason against $\neg q$, then $\neg q$ is itself a reason against one's reason. That is, 'p is a reason against $\neg q$' seems to entail '$\neg q$ is a reason against p'. From these two assumptions, we get the conclusion that 'p is a reason for q' entails '$\neg q$ is a reason against p'. As long as Contraposition holds, this conclusion makes perfect sense: If 'p is a

[12] The modal version of the account is developed in Crupi & Iacona (2020). Its probabilistic version is developed in Crupi & Iacona (2022c) and in Crupi & Iacona (2021).

[13] In this work, we will restrict consideration to sentences of the form 'p supports q', where p and q are simple categorical statements or propositional compounds thereof, not themselves including modals or other intensional operators. With this restriction, the claim that statements about reasons are contrapositive is independent from certain potentially problematic cases, such as the alleged probabilistic counterexample to modus tollens in Yalcin (2012).

reason for q' is understood as $p \rhd q$ and '$\neg q$ is a reason against p' is understood as $\neg q \rhd \neg p$, then Contraposition yields that $p \rhd q$ entails $\neg q \rhd \neg p$.[14]

Contraposition also has significant implications that involve other initially plausible principles about reasons. Here, we will focus on *Abelard's First Principle* and *Aristotle's Second Thesis*:

AFP: It is not the case that $p \rhd q$ and $p \rhd \neg q$.
AST: It is not the case that $p \rhd q$ and $\neg p \rhd q$.

Both principles have some intuitive appeal. Consider (1). If the presence of white smoke is a reason for thinking that a new pope has been chosen, then the same piece of evidence cannot also be a reason for thinking that a new pope has not been chosen. If it could, then the very notion of reason would be irremediably trivial. Similarly, if the presence of white smoke is a reason for thinking that a new pope has been chosen, then it is hard to see how the absence of white smoke can also be a reason for thinking that a new pope has been chosen. Again, this would trivialize the very notion of reason. Of course, the decision about the new pope could be reached without signalling it with white smoke, but in that case, the reason for thinking that a new pope has been chosen would *not* be the absence of white smoke.[15]

The interesting fact about Contraposition is that it makes Abelard's First Principle and Aristotle's Second Thesis interderivable. Given Substitution of Equivalents, from Contraposition we get that $p \rhd q$ and $\neg q \rhd \neg p$ are intersubstitutable, since $\neg\neg p$ and $\neg\neg q$ are equivalent to p and q. Accordingly, one can derive Aristotle's Second Thesis from Abelard's First Principle as follows:

	1	$\neg((\neg p \rhd q) \wedge (\neg p \rhd \neg q))$	AFP
A1	2	$\neg((\neg q \rhd \neg\neg p) \wedge (\neg\neg q \rhd \neg\neg p))$	SE 1
	3	$\neg((\neg q \rhd p) \wedge (q \rhd p))$	SE 2

Similarly, one can derive Abelard's First Principle from Aristotle's Second Thesis by reasoning in the opposite direction. Thus, Contraposition provides

[14] Another way to look at this point is to notice that 'being a reason against' seems a symmetric relation, unlike 'being a reason for'. (For instance, that Paul is a professional basketball player is a reason to think that he is more than 1.70 tall, but surely not the other way around.) A similar pattern holds in Crupi & Tentori (2013)'s probabilistic analysis of evidential support, where graded disconfirmation, but not positive confirmation, is the same from p to q and vice versa.

[15] Abelard's First Principle and Aristotle's Second Thesis have been widely discussed in the literature on connexive logic, see Wansing (2020). In reality, as we shall see, here we will only consider a restricted version of these principles, assuming that p is possible in the case of Abelard's First Principle and that q is not necessary in the case of Aristotle's Second Thesis. But arguably, the intuitive appeal of examples such as those considered is insensitive to the difference between restricted and unrestricted version.

a straightforward explanation of the fact that Abelard's First Principle and Aristotle's Second Thesis are both initially plausible.

The puzzle that we will consider depends on the fallibility assumption stated in Section I. Given the initial plausibility of Contraposition, it is natural to expect that there is a coherent interpretation of $\triangleright$ on which Contraposition holds but Monotonicity does not hold. However, things are not so easy. If one accepts *Right Weakening*, the rule according to which $p \triangleright q$ entails $p \triangleright r$ whenever r logically follows from q, then one cannot have Contraposition without Monotonicity. Here is the argument:

	1	$p \triangleright q$	A
	2	$\neg q \triangleright \neg p$	C 1
A2	3	$\neg q \triangleright (\neg p \vee \neg r)$	RW 2
	4	$\neg(\neg p \vee \neg r) \triangleright \neg\neg q$	C 3
	5	$(p \wedge r) \triangleright q$	SE 4

A2 shows that, given Right Weakening, if Contraposition holds, then Monotonicity holds as well. This means that, given Right Weakening, if one wants to preserve the non-monotonicity of $\triangleright$, then one must drop Contraposition.[16]

Most non-monotonic theories of conditionals validate Right Weakening, so they cannot retain Contraposition. This holds in particular for the first three interpretations considered. The conditional probability interpretation invalidates Contraposition because $P(\neg p|\neg q)$ can be lower than $P(q|p)$. Note that Aristotle's Second Thesis also fails because it can be the case that $P(q|p)$ and $P(q|\neg p)$ are both high, that is, q can be highly probable independently of p. The Stalnaker–Lewis interpretation is similar in this respect. In this interpretation, Contraposition fails because it can happen that q is true in the closest worlds in which p is true even though $\neg p$ is not true in the closest worlds in which $\neg q$ is true. Aristotle's Second Thesis also fails because q can be true both in the closest worlds in which p is true and in the closest worlds in which $\neg p$ is true. The belief revision interpretation yields the same results. Contraposition fails because it can happen that $q \in f(K, p)$ but $\neg p \notin f(K, \neg q)$. Moreover, Aristotle's Second Thesis fails because it can happen that that $q \in f(K, p)$ and that $q \in f(K, \neg p)$.

The difference-making interpretation, unlike the three interpretations just considered, invalidates Right Weakening. We regard this as a virtue. Although Right Weakening seems correct when one restricts attention to conclusive reasons, it looses part of its appeal when one reflects on some examples of non-conclusive reasons. Suppose that you are certain that Sophie is either

[16] Kraus, Lehmann, & Magidor (1990: 180–1) uses a reasoning along these lines for a consequence relation.

French or German, but that you do not know which. Consider the following sentence:

(11) If Sophie is French, then she can read.

Intuitively, (11) is not very compelling. The contrast between being French and being German does not seem relevant in order to assess Sophie's ability to read. So, it is quite plausible that, as long as conditionals are understood in the sense that matters here, (5) is acceptable but (11) is not: Although being French is a reason for being able to read French, it is not *ipso facto* a reason for being able to read. Since 'Sophie can read' logically follows from 'Sophie can read French', this implies that Right Weakening does not hold.

Note that saying that (11) is not acceptable in the sense that matters here is consistent with recognizing that (11) may be acceptable in some other sense. In particular, the consequent of (11) is highly credible given its antecedent. But the same goes for the following conditional, which certainly does not express a compelling statement about reasons:

(12) If you drink a beer, then there is snow on the Mont Blanc.

As long as an account of conditionals does not provide an adequate account of reasons, it can hardly be invoked to defend the plausibility of (11). The difference-making interpretation explains the intuitive difference between (5) and (11) as a failure of Right Weakening. (5) holds because, clearly, Sophie can read French in all the closest worlds in which she is French, whereas it is not the case that she can read French in all the closest worlds in which she is not French. (11), on the other hand, does not hold as long as it is assumed that Sophie can read in all the closest possible worlds in which she is not French.

In spite of this result, which neutralizes A2, the difference-making interpretation is not fully satisfactory. Although this interpretation preserves Aristotle's Second Thesis, it does not imply Contraposition, thus missing the explanatory connection between Abelard's First Principle and Aristotle's Second Thesis shown by A1. For example, on this interpretation, (4) is clearly acceptable, because a new pope has been chosen in all the closest worlds in which white smoke arises, and it is not the case that a new pope has been chosen in all the closest worlds in which white smoke does not arise. But it is not clear that (10) satisfies the second condition required.

The evidential interpretation provides the missing piece of the puzzle. This interpretation validates Contraposition because it follows from its very understanding of the incompatibility between p and $\neg q$ that $p \rhd q$ and $\neg q \rhd \neg p$ have the same acceptability conditions. So, the connection between Abelard's First Principle and Aristotle's Second Thesis is explained in accordance with A1. At the same time, the evidential interpretation invalidates Right Weakening, just like the difference-making interpretation, so it neutralizes A2. For example, (11) is not acceptable, unlike (5), because it is not the case that the closest possible worlds in which Sophie cannot read are worlds in which she is not French.

IV. CONFLICTING REASONS

The second puzzle concerns conflicting reasons, as it arises when distinct reasons support opposite conclusions. We will consider two versions of this puzzle: one is quite simple, the other is more sophisticated. Both versions are interesting in that they employ different principles.

The first version, considered for instance by Broome, goes as follows:[17] Apparently, it can be the case that p is a reason for q, that r is a reason for $\neg q$, and that both p and r hold. But if $\rhd$ obeys the classical rule of *Modus Ponens*, in such a case q and $\neg q$ must hold as well, which is absurd. More formally,

	1	$p \rhd q$	A
	2	$r \rhd \neg q$	A
A3	3	p	A
	4	r	A
	5	q	1,3 MP
	6	$\neg q$	2,4 MP

There are at least two ways to deal this puzzle. One is to deny that Modus Ponens holds for $\rhd$, and therefore claim that lines 5 and 6 are not justified. A plausible rationale for rejecting Modus Ponens is the observation made in Section I that reasons are non-factive, for it might be argued that the non-factivity of reasons consists precisely in the the existence of unlucky cases in which p and $p \rhd q$ are true but q is false.[18] The other is to accept A3 but argue that one of the assumptions 1–4 must be false: It cannot happen that p and r both hold and support, respectively, q and $\neg q$. In this case, it might be argued that, even though 1–4 seem to hold, at least one of them fails in some relevant sense.[19]

Each of the two options is compatible with different interpretations of $\rhd$. Consider the first option. Although Modus Ponens holds in the standard formulation of the theories of conditionals considered so far, nothing prevents us from thinking that at least some of the interpretations based on them—surely the Stalnaler–Lewis interpretation, the difference-making interpretation, and

[17] Broome (1999).

[18] Considerations along these lines can be found in Kolodny & MacFarlane (2010) and in Douven, Elqayam, & Krzyzanowska (2021).

[19] We would like to thank an anonymous referee for drawing our attention to this formulation of the puzzle.

the evidential intepretation—can be framed in terms of semantic assumptions that do not warrant this rule. Consider the second option. If one thinks that Modus Ponens must be valid, then one might still explain the failure of 1–4 by appealing to some plausible semantic story. Or at least, this route—if feasible—seems compatible with the Stalnaker–Lewis interpretation, the difference-making interpretation, and the evidential interpretation.

The second version of the puzzle employs the supposition that p and r, respectively, support q and $\neg q$ to obtain an undesirable conclusion s by means of additional principles. This version has been discussed by Nair and Horty in connection with practical reasons. Let us start with two principles that may be regarded as *prima facie* plausible when considered in isolation, *Single Reason Closure* and *Consistent Reason Agglomeration*:

SRC: If an agent has reasons to do X and X entails Y, then the agent has reasons to do Y.

CRA: If an agent has reasons to do X and has also reasons to do Y, where X and Y are consistent, then the agent has reasons to do X and Y.

Nair and Horty observe that these two principles yield weird results when applied to conflicting reasons. Imagine 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. Given Single Reason Closure, if you have a reason to meet Sam for a drink, then you have a reason to meet Sam for a drink or throw him into a canal. Given Consistent Reason Agglomeration, since you have reasons not to meet Sam for a drink, you have reasons to meet Sam for a drink or throw him into a canal and not to meet Sam. By Single Reason Closure, again, it follows that you have reasons to throw Sam into a canal.[20]

It is easy to see how the two principles considered can be converted into epistemic principles that yield similar results. Imagine that you have reasons to believe that Sam loves you, but also that you have reasons to believe that he does not love you. By the epistemic analogue of Single Reason Closure, you have reasons to believe that either Sam loves you or there are aliens. By the epistemic analogue of Consistent Reason Agglomeration, you have reasons to believe that either Sam loves you or there are aliens, and that Sam does not love you. By the epistemic analogue of Single Reason Closure, it then follows that you have reasons to believe that there are aliens.

The problem arises from the combination of three well-known principles of conditional logic. One is Right Weakening, the principle discussed in Section III. This principle captures the idea of Single Reason Closure. The other two are *Rational Monotonicity* and *AND*:

RM: $p \rhd q$ and $\neg(p \rhd \neg r)$ entail $(p \wedge r) \rhd q$.

[20] Nair (2016) and Nair & Horty (2018).

AND: $p \rhd q$ and $p \rhd r$ entail $p \rhd (q \wedge r)$.

These two principles provide a reasonable epistemic version of Consistent Reason Agglomeration. Suppose that p supports q, that r supports s, and that p and r are independent in the sense that neither of them is a reason for denying the other. Then one can derive the conclusion that p and r, taken together, constitute a reason for q and s.

A4	1	$p \rhd q$	A
	2	$r \rhd s$	A
	3	$\neg(p \rhd \neg r)$	A
	4	$\neg(r \rhd \neg p)$	A
	5	$(p \wedge r) \rhd q$	1,3 RM
	6	$(p \wedge r) \rhd s$	2,4 RM
	7	$(p \wedge r) \rhd (q \wedge s)$	5,6 AND

Thus, we get the undesirable result. Let p and r be independent reasons for q and $\neg q$, respectively. By Right Weakening, from $p \rhd q$, we get $p \rhd (q \vee s)$. From this and $r \rhd \neg q$, given A4, we get $(p \wedge r) \rhd ((q \vee s) \wedge \neg q)$. By Right Weakening again, we get $(p \wedge r) \rhd s$.[21]

A4 shows why any interpretation of $\rhd$ that validates Right Weakening, Rational Monotonicity, and AND runs into the problem illustrated. This happens in particular with the conditional probability interpretation, the Stalnaker–Lewis interpretation, and the belief revision interpretation, for each of them validates Right Weakening, Rational Monotonicity, and AND. The proofs of these facts are well established in the literature. By contrast, the difference-making interpretation and the evidential interpretation do not run into this problem because they do not validate Right Weakening.[22]

From what has been said so far about the two versions of the puzzle about conflicting reasons, it seems that at least two interpretations of $\rhd$ behave equally well with respect to this puzzle, that is, the difference-making interpretation and the evidential interpretation. As we shall see in the next section, the key difference between them emerges with the third puzzle.

V. SUPEREROGATORY REASONS

Imagine a fancy pastry shop that is renowned for its high quality. The Sachertorte—the Viennese chocolate cake—is taken to be the speciality of

[21] Note that A4 proves something stronger than Consistent Reason Agglomeration, as it does not require the consistency between X and Y. If Rational Monotonicity and AND hold, then one can prove directly $(p \wedge r) \rhd (q \wedge \neg q)$, which entails $(p \wedge r) \rhd s$ by Right Weakening.

[22] If instead one wants to retain Right Weakening, as suggested in Nair (2016), then one has to reject the conjunction of Rational Monotonicity and AND.

the place. Many people have tried it, and virtually all of them have found it exquisite. This happens to be the case not only for chocolate enthusiasts but also for those who are not generally fond of chocolate. Sally is now about to pick up a slice of Sachertorte, and she may or may not be fond of chocolate. Sally's two possible inclinations about chocolate do not have the same status relative to her attitude towards the Sachertorte. In our terms, being fond of chocolate surely is a reason to expect that she will like the cake. Not being fond of chocolate is not a reason in the same sense: Under this assumption, she will probably like the cake for its exceptionally good flavour despite her not being fond of chocolate. So, we expect that Sally will like the cake anyway, and we still recognize that in a relevant sense being fond of chocolate is a reason in support of that expectation, while not being fond of chocolate is not.

In the situation just described, it is plausible to describe Sally's being fond of chocolate as a *supererogatory reason*: p is a supererogatory reason for q when p is a reason for q but p is not needed for q to hold. Both the term and the idea of supererogatory reasons can be found in Spohn.[23] More precisely, we will say that p is a supererogatory reason for q just in case (i) $p \triangleright q$ (Sally's being fond of chocolate is a reason to think that she will like the cake), (ii) $\neg(\neg p \triangleright q)$ (Sally's not being fond of chocolate is not a reason to think that she will like the cake), and (iii) q holds whether or not p holds (Sally will like the cake whether or not she is fond of chocolate).

Not only is the notion of a supererogatory reason clearly intelligible, but it also seems able to play an independent explanatory role. In particular, when p is a supererogatory reason for q, it is typically straightforward to accept a concessive conditional that has $\neg p$ as antecedent and q as consequent. For instance, the following sentence is clearly acceptable in the case of Sally:

(13) Even if Sally is not fond of chocolate, she will like this Sachertorte.

This connection between supererogatory reasons and concessive conditionals deserves attention, or so we believe. More generally, accounting for the idea that reasons can be supererogatory in the sense just explained is a challenge for any theory of reasons.

The first four interpretations of $\triangleright$ discussed above, however, are unable to do it. In the conditional probability interpretation, (ii) and (iii) turn out to be inconsistent. If q holds whether or not p holds, then $P(q|p)$ and $P(q|\neg p)$ are both high. As a consequence, $\neg p \triangleright q$ must hold. The Stalnaker–Lewis interpretation faces a similar difficulty. If q holds whether or not p holds, then q is true in the closest world(s) in which p is true and q is also true in the closest world(s) in which p is true. But then again $\neg p \triangleright q$ holds. The same goes for the belief

[23] Spohn (2015).

revision interpretation: if q holds whether or not p holds, then $q \in f(K, p)$ and $q \in f(K, \neg p)$. As a consequence, $\neg p \rhd q$ must hold. The difference-making interpretation fails for a different reason. If q holds whether or not p holds, then q is true in the closest worlds in which p is true and q is also true in the closest worlds in which p is false. But then $p \rhd q$ turns out to be false. This makes (i) and (iii) incompatible.

The evidential interpretation, by contrast, can accommodate superogatory reasons. (i) and (ii) are obviously consistent because (i) entails (ii) in virtue or Aristotle's Second Thesis, no matter whether p is supererogatory. Moreover, the conjunction of (i) and (ii) is consistent with (iii) as long as (iii) is understood as follows: q is true both in the closest worlds in which p is true and in the closest worlds in which p is false. In particular, (iii) does not rule out (i) because it leaves room for the possibility that the closest worlds in which q is false are worlds in which p is false. This possibility, which marks a crucial distinction with respect to the difference-making interpretation, can be illustrated by supposing that, on a scale of increasing distance from the actual world, the truth values of p and q are distributed in the following order: 11,01,00,10. In this case, the conditions required by the evidential account are satisfied, hence $p \rhd q$ is true. At the same time, q is true both in the closest worlds where p is true and in the closest worlds where p is false.

A final remark concerns concessive conditionals. Arguably, a concessive conditional that has $\neg p$ as antecedent and q as consequent says that q is very likely given $\neg p$ but not in virtue of $\neg p$, because it is p, rather than $\neg p$, that supports q. For example, (13) says that it is very likely that Sally will like the Sachertorte, but not in virtue of not being fond of chocolate, because being fond of chocolate, rather than not being fond of chocolate, provides a reason for thinking that she will like the Sachertorte. As long as this analysis is granted, the evidential interpretation provides a straighforward account of the connection between superogatory reasons and concessive conditionals: p is a superogatory reason for q just in case the concessive conditional that has $\neg p$ as antecedent and q as consequent is acceptable, for the latter entails that q holds no matter whether p holds.

VI. THE LOGIC OF THE TRIANGLE

Sections III–V show how some challenging puzzles about reasons can be solved if $\rhd$ is understood in accordance with the evidential interpretation, that is, if it is assumed that $p \rhd q$ is acceptable just in case p and $\neg q$ are incompatible in the sense suggested. This section provides a more precise characterization of the evidential interpretation and explains how the logical properties of $\rhd$ can be elucidated in a suitable formal framework. In order to do so, it will be convenient to define a language that includes two symbols whose interpretation is familiar: the necessity operator $\Box$ and the symbol $>$, which represents the Stalnaker–Lewis reading of conditionals.

Let L be a language whose alphabet is constituted by a set of sentence letters p, q, r, ..., the connectives $\neg$, $\supset$, $\Box$, $>$, $\rhd$, and the brackets (,). The formulas of L are defined as follows: the sentence letters are atomic formulas; if α is a formula, then $\neg\alpha$ is a formula; if α and β are formulas, then $\alpha \supset \beta$ is a formula; if α and β are formulas which only contain the connectives just mentioned, then $\Box\alpha$, $\alpha > \beta$, $\alpha \rhd \beta$ are formulas. The connectives $\wedge$, $\vee$, $\Diamond$ are definable in terms of $\neg$, $\supset$, $\Box$, as usual.

The fact that our syntax does not allow embeddings of $\Box$, $>$, $\rhd$ makes sense given our main theoretical goal. Although embeddings would be both technically feasible and conceptually legitimate as possible extensions of L, they go beyond the purposes of a basic logic of reasons.[24] In particular, here we will make no attempt to deal with the question whether p's being a reason for q can itself be a reason for r, or if p can be a reason for q's being a reason for r.

A modal semantics for L can be defined by stipulating that every world is associated with a weak ordering of the worlds. The following definition holds for every non-empty set of worlds W:

Definition 1. *A proximity ordering O on W is an assignment to every $w \in W$ of a binary relation $\preceq_w$ that is transitive and strongly connected, that is,*
(i) for every $w', w'', w'' \in W$, if $w' \preceq_w w''$ and $w'' \preceq_w w'''$, then $w' \preceq_w w'''$;
(ii) for every $w', w'' \in W$, either $w' \preceq_w w''$ or $w'' \preceq_w w'$.

Informally speaking, $w' \preceq_w w''$ means that, from the point of view of w, w' is at least as close as w''. Accordingly, $w' \not\preceq_w w''$ means that, from the point of view of w, it is not the case that w' is at least as close as w'', that is, w'' is strictly closer than w'. The characterization of $\preceq_w$ just provided could be supplemented by adding further constraints. For example, in addition to (i) and (ii), one might require that w is w-minimal, which means that $w \preceq_w w'$ for every w'. But since (i) and (ii) will suffice for our purposes, there is no need to consider such constraints here.[25]

A model of L is defined as follows:

Definition 2. *A model of L is a triple $\langle W, O, V\rangle$, where W is a non-empty set, O is a proximity ordering on W, and V is a function such that, for each atomic formula α of L and each $w \in W$, $V(\alpha, w) \in \{1, 0\}$.*

[24] The language considered in Crupi & Iacona (2020) does allow for embeddings.

[25] Crupi & Iacona (2020) employ centred systems of spheres. Since centring entails w-minimality, this is to say that definition 1 is weaker than the stipulation about spheres adopted there.

The truth conditions of a formula α in a world w in a model M are defined as follows, where $[\alpha]_{M,w}$ indicates the value—1 or 0—that α takes in M relative to w:

Definition 3.

1. *If α is atomic, $[\alpha]_{M,w} = 1$ iff $V(\alpha,w) = 1$;*
2. $[\neg\alpha]_{M,w} = 1$ *iff* $[\alpha]_{M,w} = 0$;
3. $[\alpha \supset \beta]_{M,w} = 1$ *iff either* $[\alpha]_{M,w} = 0$ *or* $[\beta]_{M,w} = 1$;
4. $[\Box\alpha]_{M,w} = 1$ *iff, for every* w', $[\alpha]_{M,w'} = 1$.
5. $[\alpha > \beta]_{M,w} = 1$ *iff the following condition holds: if* $[\alpha]_{M,w'} = 1$ *for some* w', *then, for every* w'' *such that* $[\alpha]_{M,w''} = 1$ *and* $[\beta]_{M,w''} = 0$, *there is a* w''' *such that* $w'' \npreceq_w w'''$ *and* $[\alpha]_{M,w'''} = [\beta]_{M,w'''} = 1$;
6. $[\alpha \rhd \beta]_{M,w} = 1$ *iff the following condition holds: if* $[\alpha]_{M,w'} = 1$ *and* $[\beta]_{w'} = 0$ *for some* w', *then*
   (a) *some* w*-minimal* w'' *is such that* $[\alpha]_{M,w''} = [\beta]_{M,w''}$,
   (b) *for every* w'' *such that* $[\alpha]_{M,w''} = 1$ *and* $[\beta]_{M,w''} = 0$, *there is a* w''' *such that* $w'' \npreceq_w w'''$ *and* $[\alpha]_{M,w'''} = [\beta]_{M,w'''} = 1$, *and there is a* w'''' *such that* $w'' \npreceq_w w''''$ *and* $[\alpha]_{M,w''''} = [\beta]_{M,w''''} = 0$.

Clauses 1–4 are standard. Clause 5 specifies the truth conditions of $>$ in accordance with the Stalnaker–Lewis view: Unless α is impossible, in which case $\alpha > \beta$ is vacuously true, there must be some world in which α and β are both true and which is strictly closer than any world in which α is true and β is false. Clause 6 specifies the meaning of $\rhd$ in accordance with the evidential account: Unless it is impossible that α is true and β is false, the conjunction of (a) and (b) must hold. (a) rules out that the closest worlds are all such that α and $\neg\beta$ have the same truth value, which is a minimal condition for their being incompatibile. (b) requires that the worlds in which α is true and β is false are more distant than those in which α and β are both true or both false. It is easy to see that, due to (b), $\alpha \rhd \beta$ entails the conjunction of $\alpha > \beta$ and $\neg\beta > \neg\alpha$, which correspond, respectively, to the conditions (i) and (ii) informally stated in Section II.[26]

In definition 3, the truth conditions of a formula are specified relative to a world w in model M, that is, relative to a model-world pair M, w. This relativity amounts to relativity to an ordered set of worlds. Since the proximity ordering in M assigns a binary relation to w, namely, $\preceq_w$, to say that α is true in M, w is to say that α is true relative to $\preceq_w$.

Logical consequence is defined as follows, for any set of formulas Γ and every formula α.

[26] This formulation slightly differs from the truth condition for the evidential conditional originally provided in Crupi & Iacona (2020). The difference is explained in Crupi & Iacona (2022a).

Definition 4. $\Gamma \models \alpha$ *iff, for every M, w, if all the formulas in* Γ *are true in M, w, then* α *is true in M, w.*

This semantics yields some results about $\triangleright$ that are reasonable to expect. As Crupi and Iacona have shown, the following facts are provable:

Fact 1. $\alpha \triangleright \gamma \not\models (\alpha \wedge \beta) \triangleright \gamma$.

Fact 2. $\alpha \triangleright \beta \models \neg\beta \triangleright \neg\alpha$.

Fact 3. $\Diamond\alpha \models \neg((\alpha \triangleright \beta) \wedge (\alpha \triangleright \neg\beta))$.

Fact 4. $\Diamond\neg\beta \models \neg((\alpha \triangleright \beta) \wedge (\neg\alpha \triangleright \beta))$.

Fact 5. *Not: if* $\beta \models \gamma$, *then* $\alpha \triangleright \beta \models \alpha \triangleright \gamma$.

Fact 6. *If* $\alpha \models \beta$, *then* $\models \alpha \triangleright \beta$.

Fact 7. $(\alpha \triangleright \beta) \wedge (\alpha \triangleright \gamma) \models \alpha \triangleright (\beta \wedge \gamma)$.

Fact 8. $\Box(\alpha \supset \beta) \models \alpha \triangleright \beta$.

Fact 1 shows that Monotonicity does not hold for $\triangleright$. Facts 2–4 show that Contraposition, a restricted version of Abelard's First Principle, and a restricted version of Aristotle's Second Thesis hold for $\triangleright$. Fact 5 shows that Right Weakening does not hold for $\triangleright$. Fact 6 expresses the principle known as *Supraclassicality*, which accords with what has been said in section I about the connection between reasons and inferences: α supports β when β logically follows from α. Fact 7 expresses AND, which is also plausible: If α supports both β and γ, then it supports $\beta \wedge \gamma$. Finally, fact 8 shows that the strict conditional is stronger than the evidential conditional: If α necessitates β, then α supports β. Since the strict conditional is monotonic, this is to say that conclusive reasons are a proper subclass of reasons.[27]

Some of the facts just stated are directly relevant to the three puzzles discussed in Sections III–VI. In particular, the first puzzle hinges on Contraposition, Abelard's First Principle and Aristotle's Second Thesis. The second puzzle involves AND and Right Weakening (along with Rational Monotonicity, which is also not provable in our semantics). Finally, in the discussion of the third puzzle, we observed that (i) entails (ii) in virtue of Aristotle's Second Thesis. The other observation, namely, that (i) and (ii) do not rule out (iii), can be proved by costructing a model in which there are four worlds that display a distribution of truth values for α and β of the kind suggested in Section V, that is, a distribution which characterizes α as a supererogatory reason for β.

As anticipated in Section II, the modal semantics just outlined is not the only way to spell out the incompatibility condition that characterizes the

[27] Crupi & Iacona (2020). As it is easy to verify, the stronger semantic assumptions made in that work (see previous footnote) are not essential to the proof of facts 1–8. Raidl, Iacona, & Crupi (2021) provide a sound and complete axiom system called EC, which displays the distinctive logical properties of the interpretation of $\triangleright$ obtained with the stronger semantics.

evidential interpretation. As we have seen, there is a coherent alternative to it that adopts a probabilistic measure. For any $\alpha, \beta \in \mathsf{L}$ and any probability function P defined over L, the degree of incompatibility between α and $\neg\beta$ can be represented in terms of that measure. The acceptability of $\alpha \rhd \beta$—which indicates the strength of α as a reason for β—can then be equated with the degree of incompatibility between α and $\neg\beta$. In other words, the semantics of L can be given in terms of a function A defined for any probability function P in such a way that the following holds:

Definition 5.

$$A_P(\alpha \rhd \beta) = \begin{cases} 1 - \frac{P(\alpha \wedge \neg\beta)}{P(\alpha)P(\neg\beta)} & \textit{if } P(\alpha \wedge \neg\beta) \leq P(\alpha)P(\neg\beta), \\ 1 & \textit{if } P(\alpha) = 0 \textit{ or } P(\beta) = 1, \\ 0 & \textit{otherwise}. \end{cases}$$

For any formula α of L, $A_P(\alpha)$ represents the degree of acceptability of α given P. Validity can be defined in terms of degree of uncertainty in the way suggested by Adams, assuming that the uncertainty of α given P is expressible as $1 - A_P(\alpha)$. As shown by Crupi and Iacona, the resulting logic largely converges with the results obtained with the modal semantics above. In particular, facts 1–8 turn out to be provable.[28]

VII. SUFFICIENT REASONS, NECESSARY REASONS, AND DIFFERENCE-MAKING

One way to illustrate the explanatory potential of the formal account outlined in the previous section is to show how some main notions that are currently employed in discussions about reasons can be expressed in L. Let us start with the notion of *sufficient reason*, which is widely adopted in everyday language. As it emerges from the initial clarifications given in Section I, we understand reasons as sufficient reasons: to say that p is a reason for q is to say that p provides a justification for believing q, so that q can be inferred from p. For example, Sophie's being French is sufficient for thinking that she can read French. More generally, the following equivalence holds:

Definition 6. *p is a sufficient reason for q iff $p \rhd q$.*

[28] Crupi & Iacona (2022c). According to Definition 5, the acceptability of $\alpha \rhd \beta$, hence the strength of α as a reason for β, amounts to the degree of evidential support from α to β as characterized in Crupi & Tentori (2013). Alternative probabilistic accounts of the strength of reasons have been recently discussed by Kernohan (2022) and Nair (2021).

The right-hand side of this biconditional is to be understood as an adequate formalization in L of its left-hand side.[29]

It is important to note that sufficiency, so understood, is consistent with fallibility. One way to see this is to recall fact 1: since $\triangleright$ is non-monotonic, the right-hand side of definition 6 says at most that p is a defeasible reason for q. So, 'sufficient' can be read as 'defeasibly sufficient'. The claim that p is a conclusively sufficient reason for q is expressed by the formula $\Box(p \supset q)$, which entails $p \triangleright q$ by fact 8. For example, Sophie's capacity to read French is conclusively sufficient for thinking that she can read.

Another way to render fallibility is to say that sufficiency does not entail factivity: Having a sufficient reason for q does not guarantee that q is true. As noted in Section IV, one might equate factivity with Modus Ponens and argue that $\triangleright$ must not obey Modus Ponens. Alternatively, one might provide some account of non-factivity that is consistent with Modus Ponens. Both options are compatible with the semantics provided in section VI, given that definition 1 does not include w-minimality, the condition that warrants Modus Ponens. Since for the purposes of this paper there is no need to choose between them, we will remain neutral here as to the question whether Modus Ponens holds for $\triangleright$.

Sufficient reasons are often contrasted with *necessary reasons*. To say that p is a necessary reason for q is to say that, in order to believe q, one has to accept p. Or equivalently, not accepting p prevents one from believing q. For example, one can hardly believe that Sophie is French without also believing that she can read French. The claim that p is a necessary reason for q can be expressed in L by the formula $q \triangleright p$, which is equivalent to $\neg p \triangleright \neg q$. If one rejects p, then one has a reason for rejecting q.

Definition 7. *p is a necessary reason for q iff $q \triangleright p$.*

As in the case of sufficient reasons, necessary reasons can be fallible. In particular, 'necessary' can be read as 'defeasibly necessary'. The claim that p is a conclusively necessary reason for q is expressed by the formula $\Box(q \supset p)$, which entails $q \triangleright p$. For example, Sophie's capacity to read is conclusively necessary for thinking that she can read French.

Given definitions 6 and 7, the stronger claim that p is a necessary and sufficient reason for q is expressed in L by a conjunction of formulas:

Definition 8. *p is a necessary and sufficient reason for q iff $(p \triangleright q) \wedge (q \triangleright p)$.*

An example of a necessary and sufficient reason that can be represented in this way is the following: The presence of smoke is necessary and sufficient for

[29] This is not to deny that an intelligible notion of *in*sufficient reason can be defined. Broome (1999) and Spohn (2012), in different ways, contemplate such a notion, as a reason that positively contributes to credibility but is not quite enough for inference. Our probabilistic semantics would also allow for a formal analysis of this idea.

thinking that there is a fire. The treatment of conclusive reasons is exactly as in the previous two cases.

The formal characterization of sufficient and necessary reasons just outlined accords with three plausible assumptions, which are often made about sufficient and necessary conditions in general. The first is that p is sufficient for q if and only if q is necessary for p: this holds simply because the same formula $p \rhd q$ expresses both the sufficiency of p for q and the necessity of q for p. The second is that q is necessary for p if and only if $\neg q$ is sufficient for $\neg p$: this holds because $p \rhd q$ is equivalent to $\neg q \rhd \neg p$. The third is that necessary and sufficient conditions are symmetrical, that is, p is necessary and sufficient for q if and only if q is necessary and sufficient for p: this holds in virtue of the symmetry of the formula $(p \rhd q) \wedge (q \rhd p)$.

Lastly, we will consider the notion of *difference-making*, which is often associated with the idea of support. Intuitively, to say that p makes a difference for q is to say that q is credible on the assumption that p but not otherwise. For example, we can easily imagine a case in which a patient shows a highly distinctive combination of symptoms that indicate a certain disease, but it would be wrong to diagnose the same disease if those symptoms were absent. As it emerges from Section V, we do not regard difference-making as an essential feature of reasons: The case of supererogatory reasons shows that p can be a reason for q even though p makes no difference for q in the sense specified. Nonetheless, difference-making can be expressed in L as an additional condition for reasons. More precisely, supererogatory and difference-making reasons can be characterized as mutually exclusive categories of reasons individuated in terms of such a condition.[30]

Let us start with supererogatory reasons. As noted in Section V, it is plausible to think that p is a supererogatory reason for q when a concessive conditional that has $\neg p$ as antecedent and q as consequent is acceptable. Arguably, the logical form of the latter conditional can be represented as follows: $(\neg p > q) \wedge (p \rhd q)$.[31] Inverting the order of the two conjuncts, this means that p is a reason for q, but q is also credible on the assumption that $\neg p$.

Definition 9. *p is a supererogatory reason for q iff* $(p \rhd q) \wedge (\neg p > q)$.

For example, Sally's being fond of chocolate is a reason for thinking that she will like the Sachertorte, but she will probably like the Sachertorte even if she is not fond of chocolate. Being fond of chocolate is not necessary in this respect. Note that definitions 7 and 9 entail that, if p is a supererogatory reason for q, then p is not a necessary reason for q, given that $\neg p > q$ rules out that the

[30] By taking sufficient reasons to be partitioned into supererogatory and diference-making, we depart from the terminology employed in Spohn (2012), and which goes back to Spohn (1991). Spohn identifies Rott's definition of difference-making with 'being a sufficient reason', so that in his framework, sufficient and supererogatory turn out to denote incompatible features of reasons.

[31] This is the analysis of concessive conditionals suggested in Crupi & Iacona (2022b).

closest worlds in which $\neg p$ holds are worlds in which $\neg q$ holds, which instead is required by $q \rhd p$.

Difference-making reasons are the opposite of supererogatory reasons in the following sense: p is a difference-making reason for q when p is a reason for q and it is *not* the case that q is credible on the assumption that $\neg p$. So, in this case, the additional conjunct is $\neg(\neg p > q)$ rather than $\neg p > q$.

Definition 10. *p is a difference-making reason for q iff* $(p \rhd q) \wedge \neg(\neg p > q)$.

Definitions 9 and 10 imply that being supererogatory and being difference-making are mutually exclusive and jointly exhaustive properties of reasons: For any p and q such that $p \rhd q$, exactly one of the formulas $\neg p > q$ and $\neg(\neg p > q)$ must be true.

Note that from definitions 8 and 10 we get that, as long as $\neg p$ is possible, if p is a necessary and sufficient reason for q, then p is a difference-making reason for q. This holds because the second conjunct of $(p \rhd q) \wedge (q \rhd p)$ entails $\neg p \rhd \neg q$ by fact 2, and the latter entails $\neg(\neg p \rhd \neg\neg q)$ by fact 3, which is equivalent to $\neg(\neg p > q)$. For example, since the presence of smoke is necessary and sufficient for thinking that there is a fire, it is also a difference-making reason in the sense defined.

VIII. CONDITIONALS AND ARGUMENTS

So far, we have employed the symbol $\rhd$ to outline an analysis of 'p is a reason for q'. This symbol behaves as a conditional, so its logical properties are naturally framed as principles of conditional logic. But the core idea of the evidential interpretation can also be applied to arguments, as noted in Section I, for whenever p is a reason for q, the inference from p to q is justified. As long as validity is understood as the property that an argument has when its conclusion is justifiedly inferred from its premises, we get that whenever $p \rhd q$ holds, the argument from p to q is valid. So, the evidential interpretation implies that there is a straightforward sense in which valid arguments are equivalent to true conditionals.[32]

In order to provide a more precise characterization of this equivalence, it must be taken into account that validity exhibits one distinctive form of relativity. Consider, for example, (7). To a first approximation, to say that the inference stated in (7) is justified is to say that, given some body of evidence that constitutes our background information—the death of the former pope, the electoral procedures of the College of Cardinals, and so on—from the premise that white smoke raises, it is plausible to conclude that a new pope

[32] Iacona (forthcoming) provides a detailed discussion of the thesis that valid arguments amount to true conditionals and spells out some of its implications.

has been chosen. The qualification 'given some body of evidence' is necessary here because it indicates the epistemic context in which the argument is assessed. Clearly, the same inference would not be regarded as justified if the circumstances were relevantly different, say, if the pope were still alive or the Vatican had entirely different laws. More generally, arguments are judged valid or invalid relative to sets of background assumptions that hold in the context in which they are used.

We will call *circumstances of evaluation* this parameter. Circumstances of evaluation are to be understood as sets of assumptions that can warrant the inference from a set of premises to a conclusion. These assumptions concern not only information about what is actually the case but also information about what is necessarily the case or what is likely to be the case. There are at least two ways to represent circumstances of evaluation, which correspond to the two versions of the evidential interpretation. One is to take circumstances of evaluation to be ordered sets of worlds. The other is to take them as probability assignments. In both cases, the index c will be used to refer to circumstances of evaluation, and the index symbol $\Rightarrow_c$ will indicate validity in c.

The equivalence between arguments and conditionals may thus be phrased as follows, where α is any conjunction of formulas of L:

Definition 11. *$\alpha \Rightarrow_c \beta$ iff $\alpha \rhd \beta$ is acceptable in c.*

If circumstances of evaluation are identified with ordered sets of words, then c can be replaced by $\langle M, w\rangle$. For example, the argument stated in (7) is valid relative to a world where the former pope is dead, where the College of Cardinals follows certain electoral procedures, where the closest worlds in which white smoke raises are worlds in which a new pope has been chosen, and so on.

Instead, if circumstances of evaluation are identified with probability assignments, and 'acceptable' is understood as 'having a degree of acceptability above a given threshold', then c can be replaced by P, and the right-hand side of definition 11 can be understood in terms of the probabilistic measure specified in definition 5. Without such a threshold, the definition can simply be rephrased as equating a degree of relative validity with a degree of relative acceptability. In the latter case, relative validity is treated as a gradable property of arguments, which is in line with the widespread assumption that inductive strength allows for degrees.

No matter which of the two options is adopted, or whether some other route is pursued to model contexts, there are 'absolute' facts about validity that can be derived from the relative notion of validity by quantifying over circumstances of evaluation. These absolute facts can be regarded as properties of a non-monotonic consequence relation $|\!\sim$ that characterizes reasons in general. We will provide two examples to illustrate this point. First, consider

Supraclassicality. In virtue of fact 6, we have that if $\alpha \models \beta$, then $\alpha \Rightarrow_c \beta$ for every c. This can be rewritten as follows:

Fact 9. *If* $\alpha \models \beta$, *then* $\alpha \mid\sim \beta$.

Proof. From fact 6 and definition 11. □

Second, consider Contraposition. In virtue of fact 2, we have that, for every c, if $\alpha \Rightarrow_c \beta$, then $\neg\beta \Rightarrow_c \neg\alpha$. Again, we can use the symbol $\mid\sim$ to express this fact:

Fact 10. *If* $\alpha \mid\sim \beta$, *then* $\neg\beta \mid\sim \neg\alpha$.

Proof. From fact 2 and definition 11. □

The relationships between systems of conditional logic and formal accounts of non-monotonic consequence relations have been extensively addressed, and different methods have been explored to establish fruitful connections between them.[33] According to the line of thought suggested here, which hinges on the equivalence between valid arguments and true conditionals, a system that captures the logical properties of $\triangleright$ and a corresponding account of $\mid\sim$ may be regarded as two distinct theories that stem from the same conceptual source.

In a related project on which we are currently working, we define a minimal logic for a consequence relation of the kind illustrated above. More precisely, the fundamental rules that define this logic include Supraclassicality and Contraposition—that is, facts 9 and 10—along with the replaceability of classically equivalent formulas for both premises and conclusion. Being both non-monotonic and contrapositive, such a consequence relation demonstrably departs from the fundamental properties of non-monotonic logics as initially defined in Gabbay's seminal work and then developed in established systems such as KLM logic.[34] In our view, this departure uncovers interesting logical facts that deserve careful consideration.

IX. REASONS AND INDUCTIVE LOGIC

Inductive logic has been traditionally regarded as a generalization of deductive logic. As deductive logic is the theory of conclusive, non-defeasible arguments, inductive logic should be a theory of possibly non-conclusive, defeasible arguments; hence, it must imply a failure of Monotonicity. The question then naturally arises of which of the properties of deductive logic can instead be retained in the generalization to inductive logic. The case of Contraposition and Right Weakening is particularly interesting in this respect. When one has

[33] See the classical discussion in Makinson (2005) and the more recent survey in Strasser & Antonelli (2019).

[34] Gabbay (1985) and Kraus, Lehmann, & Magidor (1990).

a deductively valid argument from a premise α to a conclusion β, one can always safely infer two things. First, the same premise α entails any conclusion γ that is logically weaker than β. Second, the negation of the conclusion, $\neg\beta$, entails the negation of the premise, $\neg\alpha$. If Monotonicity fails, however, one cannot have both things: as we have seen, the conjunction of Contraposition and Right Weakening entails Monotonicity. A theory of inductive validity can only imply *either* Right Weakening *or* Contraposition.

This dilemma can be traced back to Carnap's project in *Logical Foundations of Probability* (1950–62). Carnap's 'system of inductive logic' in the second part of that work is based on a probabilistic measure of relevance, which shares two key properties with our own proposal as long as we equate the (positive) relevance of α to β and the degree of incompatibility of α with $\neg\beta$.[35] Carnap's measure encodes the idea of Contraposition in the sense that, for any given probability function and set of background assumptions (so in any context, in our terms), it assigns the same numerical value to the pairs of statements α, β and $\neg\beta$, $\neg\alpha$, where α increases the probability of β to some extent (and likewise for $\neg\beta$ and $\neg\alpha$). Moreover, Carnap's measure implies failure of Right Weakening because, being non-monotonic, it assigns a lower value to α, β than to γ, β in some contexts even if γ logically follows from α. Accordingly, it must assign a lower value to $\neg\beta$, $\neg\alpha$ than to $\neg\beta$, $\neg\gamma$ in those contexts even if $\neg\alpha$ logically follows from $\neg\gamma$. Thus, a strong inductive argument can be turned into a weaker one by weakening the conclusion, because the positive relevance of the premise can decrease or get lost. Our proposal is similar in both respects, in that it validates Contraposition and yields that α can be a reason for β without thereby being a reason for something logically weaker than β.

When Carnap discussed Hempel's seminal work on confirmation in the 40s (§87), he addressed the so-called Special Consequence Condition, in fact a rather precise analogue of Right Weakening in the framework of confirmation theory. According to the Special Consequence Condition as advocated by Hempel, if α confirms β, then α also confirms anything that logically follows from β. In the 1962 Preface to his masterpiece, Carnap famously untied the knot of the tension between his view and Hempel's by postulating two distinct explicata for inductive logic: the *firmness* of a statement given some evidence and the *increase in firmness* of a statement as provided by some evidence. Interpreting inductive logic as a conservative non-monotonic extension of deductive logic, Right Weakening is key for the former concept (following the Hempelian way, we might say), while Contraposition can plausibly be preserved only for the latter (following the Carnapian way).

[35] Carnap (1950/62). The two measures differ in other important respects, which do not concern us here, however, see Crupi, Tentori, & Gonzalez (2007).

Both the Carnapian and the Hempelian way have rather peacefully survived in contemporary confirmation theory.[36] The theoretical potential of the Carnapian way, however, has remained largely unexplored in other domains, such as theories of non-monotonic conditionals. In fact, three of the four interpretations of $\rhd$ that we discussed above (Adams's, the Stalnaker-Lewis, and the belief revision account) follow the Hempelian way, while our analysis of $\rhd$ is distinctive among prominent theories in that it satisfies Contraposition instead of Right Weakening. Relying on the tight connection between defeasible reasons, non-monotonic conditionals, and inductive arguments, the Carnapian way can therefore be fully articulated in a consistent manner, we submit.[37]

REFERENCES

Adams, E. W. (1965) 'The Logic of Conditionals', *Inquiry*, 8: 166–97.
Arlo-Costa, H. (1995) 'Epistemic Conditionals, Snakes and Stars', in G. C. L. Farinas del Cerro and A. Herzig (eds.) *Conditionals, from Philosophy to Computer Science*, 193–239. Oxford: OUP.
Broome, J. (1999) 'Normative Requirements', *Ratio*, 12: 398–419.
Carnap, R. (1950/62) *Logical Foundations of Probability*. Chicago, IL: University of Chicago Press.
Crupi, V. (2021) 'Confirmation', in *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2021/entries/confirmation.
Crupi, V. and Iacona, A. (2020) 'The Evidential Conditional', *Erkenntnis*, doi: 10.1007/s10670-020-00332-2.
—— (2021) 'Probability, Evidential Support, and the Logic of Conditionals', *Argumenta*, 6: 211–22.
—— (2022a) 'Conditionals: Inferentialism Articulated', manuscript.
—— (2022b) 'On the Logical Form of Concessive Conditionals', *Journal of Philosophical Logic*, 51: 633–51.
—— (2022c) 'Three Ways of Being Non-Material', *Studia Logica*, 110: 47–93.
Crupi, V. and Tentori, K. (2013) 'Confirmation as Partial Entailment: A Representation Theorem in Inductive Logic', *Journal of Applied Logic*, 11: 364–72.
Crupi, V., Tentori, K. and Gonzalez, M. (2007) 'On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues', *Philosophy of Science*, 74: 229–52.
Douven, I., Elqayam, S. and Krzyzanowska, K. (2021) 'Inferentialism: A Manifesto', in S. Kaufmann, D. Over and G. Sharma (eds.) *Conditionals: Logic, Linguistics and Psychology*. London: Palgrave MacMillan.
Faroldi, F. L. G. and Protopopescu, T. (2019) 'A Hyperintensional Logical Framework for Deontic Reasons', *Logic Journal of the IGPL*, 27: 411–33.
Fuhrmann, A. (2017) 'Why Abandon the Default Approach', *Erkenntnis*, 82: 1351–65.
Gabbay, D. M. (1985) 'Theoretical Foudation for Non-Monotonic Reasoning in Expert Systems', in K. R. Apt (ed.) *Logic and Models of Concurrent Systems*. Berlin: Springer.
Gärdenfors, P. (1988) *Knowledge in Flux*. Cambridge, MA: MIT Press.
Goble, L. (2013) 'Prima Facie Norms, Normative Conflicts, and Dilemmas', in D. M. Gabbay, J. Horty and X. Parent (eds.) *Handbook of Deontic Logic and Normative Systems*, 241–351. London: College Publications.
Horty, J. (2007) 'Reasons as Defaults', *Philosopher's Imprint*, 7: 1–28.
—— (2012) *Reasons as Defaults*. Oxford: OUP.
Iacona, A. (forthcoming) 'Valid Arguments as True Conditionals', *Mind*.

[36] See Crupi (2021).
[37] We would like to thank José Diez, Wolfgang Spohn, Jan Sprenger, and two anonymous reviewers for their helpful comments on previous versions of this paper.

Kernohan, A. (2022) 'How to Modify the Strength of a Reason', *Philosophical Studies*, 179: 1205–20.
Kolodny, N. and MacFarlane, J. (2010) 'Ifs and Oughts', *The Journal of Philosophy*, 107/3: 115–43.
Kraus, S., Lehmann, D. and Magidor, M. (1990) 'Nonmonotonic Reasoning, Preferential Models and Cumulative Logics', *Journal of Artificial Intelligence*, 44: 167–207.
Levi, I. (1988) 'Iteration of Conditionals and the Ramsey Test', *Synthese*, 76: 49–81.
Lewis, D. (1973) *Counterfactuals*. Oxford: Blackwell.
Makinson, D. (2005) *Bridges from Classical to Nonmonotonic Loigc*. London: King's College Publications.
Nair, S. (2016) 'Conflicting Reasons, Unconflicting 'Ought's', *Philosophical Studies*, 173/3: 629–63.
—— (2021) "Adding up' Reasons: Lessons for Reductive and Non-Reductive Approaches', *Ethics*, 132: 38–88.
Nair, S. and Horty, J. (2018) 'The Logic of Reasons', in D. Star (ed.) *The Oxford Handbook of Reasons and Normativity*. Oxford: OUP.
Pollock, J. L. (1970) 'The Structure of Epistemic Justification', in N. Rescher (ed.) *Studies in the Theory of Knowledge. American Philosophical Quarterly Monograph Series 4*, 62–78. Oxford: Basil Blackwell.
—— (1987) 'Defeasible Reasoning', *Cognitive Science*, 11: 481–518.
—— (1995) *Cognitive Carpentry: A Blueprint for How to Build a Person*. Cambridge, MA: MIT Press.
Raidl, E. (2021) 'Definable Conditionals', *Topoi*, 40: 87–105.
Raidl, E., Iacona, A. and Crupi, V. (2021) 'The Logic of the Evidential Conditional', *Review of Symbolic Logic*, 1–13, doi: 10.1017/S1755020321000071.
Reiter, R. (1980) 'A Logic for Default Reasoning', *Artificial Intelligence*, 13: 81–132.
Rott, H. (1986) 'Ifs, Though, and Because', *Erkenntnis*, 25: 345–70.
—— (2022) 'Difference-Making Conditionals and the Relevant Ramsey Test', *Review of Symbolic Logic*, 15: 133–64.
Spohn, W. (1991) 'A Reason for Explanation: Explanations Provide Stable Reasons', in W. Spohn, B. C. Van Fraassen and B. Skyrms (eds.) *Existence and Explanation*, 165–96. Berlin: Springer.
—— (2012) *The Laws of Belief*. Oxford: OUP.
—— (2015) 'Conditionals: A Unifying Ranking-Theoretic Perspective', *Philosophers' Imprint*, 15: 1–30.
Stalnaker, R. (1991) 'A Theory of Conditionals', in F. Jackson (ed.) *Conditionals*, 28–45. Oxford: OUP.
Strasser, C. and Antonelli, G. A. (2019) 'Non-monotonic Logic', in E. N. Zalta (ed.) *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/fall2018/entries/logic-nonmonotonic.
Wansing, H. (2020) 'Connexive Logic', in E. Zalta (ed.) *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/spr2019/entries/logic-connexive.
Yalcin, S. (2012) 'A Counterexample to Modus Tollens', *Journal of Philosophical Logic*, 41: 1001–24.

University of Turin, Italy