Brandom’s Account of Defeasible Reasoning: Problems and Solutions

by

Reiner Schaefer

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
For the degree of
Doctorate of Philosophy
In
Philosophy

Guelph, Ontario, Canada

© Reiner Schaefer, April, 2012
AbSTRACT

Brandom’s Account of Defeasible Reasoning:
Problems and Solutions

Reiner Schaefer          Advisor:  
University of Guelph, 2012  Dr. Mark McCullagh

Robert Brandom has provided what is probably one of the best worked out accounts of how the meanings of linguistic expressions are determined by how they are used—in particular, used in inferences. There are three different types of inferential relations in terms of which Brandom gives his account: commitment-preserving, entitlement-preserving, and incompatibility relations.

Brandom also recognizes that most of the reasoning we engage in is defeasible (or deductively inconclusive). For example, the inference from ‘Tweety is a bird’ to ‘Tweety can fly’ is defeasible, because it can be defeated if there is stronger overriding reason to deny that ‘Tweety can fly’—such as Tweety’s being a penguin. Surprisingly, Brandom’s three types of inferential relations are inadequate for describing defeasible inference. In my dissertation I explain how the problem arises—it’s actually two problems—and I propose a solution that is consistent with Brandom’s overall approach.

The first problem is that although Brandom's account does explain how someone can lose entitlement to a claim by committing themselves to some other claim, as in the Tweety example, it doesn’t allow subsequent recovery of entitlement to that claim by the addition of yet further information—say, that Tweety is a penguin with a jetpack. Once defeated (by some information), an inference stays defeated, on Brandom's account. The second problem is that of interpretation: when should we interpret someone as committed to the propriety of an inference that is defeasible? Brandom's account of what it is to endorse an inferential relation has no room for the important distinction between endorsing an inference in a context in which it happens to be defeated, and not endorsing it at all.

In the latter portion of this dissertation I propose various modifications to Brandom’s account that will allow it overcome these problems. I solve the first problem by modifying Brandom’s account of how someone is obliged to update their beliefs in light of the inferential relations they endorse. I solve the second problem by modifying Brandom’s account of when we can appropriately interpret someone as endorsing particular inferential relations.
Dedicated to my loving wife and best friend Julia Harrington.

Thanks for making the sacrifices that let me pursue my PhD.

Thanks for your patience with me as I wrote this.

Thanks for your moral support.
Acknowledgements

First I would like to acknowledge the indirect contributions made by three of my former professors. Prof Neil McGill’s courses (and conversations at his dinner parties) at the University of New Brunswick inspired my interest in analytic philosophy of language and mind. Prof Ausonio’s Marras’s philosophical mentorship was invaluable during my time at the University of Western Ontario. Prof Bruce Freed’s stimulating course on Brandom’s *Making It Explicit* inspired me to write my dissertation on Brandom’s work.

Next, I am grateful to the University of Guelph Philosophy Department for its intellectual, financial, and moral support. It provided an excellent environment for working through the ideas that this dissertation is built upon.

I am especially grateful to the members of my excellent advisory committee: to Andrew Bailey, Rockney Jacobsen, and particularly to my supervisor Mark McCullagh. The arguments and ideas presented in this dissertation would not be nearly as well refined or developed as they are without Prof McCullagh’s thoughtful comments on the many drafts. Thanks also go out to those who were part of the examining committee: Tim Kenyon and Peter Loptson.

This research was partially funded by the Ontario Government through their OGS scholarships.
# Contents

Overview ix

I: General Introduction to Brandom’s Account of Language and Reasoning and his Commitment to its being Predominantly Defeasible 1

1: An Outline of how Brandom Grounds Meaning in Practice 1

2: Setting the Stage 9

3: The Gap in the Secondary Literature 22

II: Brandom’s Account of the Consequences of Attributing Inferential Commitments, and Why it does not Allow Entitlement-Recovery 28

1: The General Strategy 28

2: Brandom’s Account of the Consequences of Attributing Inferential Commitments. 29

3: Why Brandom’s Account of the Consequences of Attributing Inferential Commitments does not Allow Entitlement-Recovery 36

III: Brandom’s Account of the Appropriate Circumstances for Attributing Inferential Commitments and Why it Does Not Adequately Allow for Defeaters that are not Refuters 50

1: Defeaters that are not Refuters 50

2: Informal Presentation of Brandom’s Account of the Appropriate Circumstances for Attributing Inferential Commitments. 53

3: Putting All the Pieces Together: A More Formal Presentation of Brandom’s Account of the Appropriate Circumstances for Attributing Inferential Commitments 71

4: Why Brandom’s Account of the Circumstances for Attributing Inferential Commitments does not allow for Defeaters that are not Refuters 81
5: Two Unsuccessful Quick Fixes 88

IV: Some Pessimism about the Prospects of Incorporating Features from Prominent Accounts of Defeasible Reasoning 94

1: The Challenges of Jury-rigging 94
2: John Pollock’s “Defeasible Reasoning” 99
3: Raymond Reiter’s Default Logic 107
4: AGM on Belief Revision and Nonmonotonic Inference 116

V: Solving the Problem of Entitlement-Recovery: The Revised Account of the Consequences of Attributing Commitment to the Three Different Types of Inferential Relations 131

1: Clarifying How the Expressions Commonly Employed in Discussions of Defeasible Reasoning will be Used 132
2: A Review of the Problem of Entitlement-Recovery 139
3: A Revised Account of Defeat 142
4: Revising the Scorekeeping Algorithm 158
5: Putting to Work the Revised Account of the Consequences of Inferential Commitment 166

VI: Solving the Problem of Defeaters that are not Refuters: The Revised Account of the Appropriate Circumstances for Attributing the Three Different Types of Material Inferential Relations 173

1: The Challenge at Hand 173
2: Accomodating Inferential Holism 176
3: The Revised Account of the Appropriate Circumstances for Attributing Implicit Commitments to Indefeasible Inferential Relations 184
4: The Revised Account of Appropriate Circumstances for Attributing Implicit Commitments to Defeasible Inferential Relations 187
5: Applying the Revised Account to the Match-Lighting Scenario  201
6: How the Revised Account Allows for Defeaters that are not Refuters  203

Summary and Concluding Remarks
Overview:

In his book *Making It Explicit* (and in various related works\(^1\)) Robert Brandom has provided what is probably one of the best worked out accounts of how the *meanings* of linguistic expressions are determined by how they are *used*—in particular, used in inferences. There are three different types of inferential relations in terms of which Brandom gives his account: commitment-preserving, entitlement-preserving, and incompatibility relations.

Brandom also recognizes that most of the reasoning we engage in is *defeasible* (or deductively inconclusive). For example, the inference from ‘Tweety is a bird’ to ‘Tweety can fly’ is defeasible, because it can be defeated if there is stronger overriding reason to deny that ‘Tweety can fly’—such as Tweety’s being a penguin. Surprisingly, Brandom’s three types of inferential relations are inadequate for describing defeasible inference. In my dissertation I explain how the problem arises—it’s actually two problems—and I propose a solution that is consistent with Brandom’s overall approach.

The first problem is that although Brandom’s account does explain how someone can *lose* entitlement to a claim by committing themselves to some other claim, as in the Tweety example, it doesn’t allow subsequent *recovery* of entitlement to that claim by the addition of yet further information—say, that Tweety is a penguin with a jetpack. Once defeated (by some information), an inference stays defeated, on Brandom’s account.

The second problem is that of interpretation: when should we interpret someone as committed to the propriety of an inference that is defeasible? Brandom’s account of

---

\(^1\) In particular see Brandom’s *Articulating Reasons*, and *Between Saying and Doing*
what it is to endorse an inferential relation has no room for the important distinction between endorsing an inference in a context in which it happens to be defeated, and not endorsing it at all.

Brandom is not the only theorist who has difficulty accounting for defeasible reasoning. But some of the core features of Brandom’s broader theory of language and reasoning make the problem take a different shape for him than it might for other theorists. One example of this is Brandom’s commitment to a sort of Quinean holism that blurs the distinction between what someone believes and what inferential relations that person takes different claims to stand in—or between what the facts are and what different expressions mean.² Another example is Brandom’s commitment to reasoning being thoroughly social.

I solve the problem of entitlement-recovery by modifying the way that Brandom’s three different types of inferential relations interact with one another, making it so that the premises of one defeasible inferential relation (eg *Tweety is a penguin with a jetpack → Tweety can fly*) can be a reason against the conclusion of another defeasible inferential relation (*Tweety is a penguin → Tweety cannot fly*). I then appeal to a version of the principle of specificity³ so that the inference with the more specific set of premises (eg Tweety is a penguin with a jetpack) overrules the inference with less specific premises (Tweety is a penguin).

---

² See Quine (1951), and Quine (1995)

³ The principle of specificity is commonly appealed to in the Artificial Intelligence literature on defeasible reasoning. It, very roughly, states that if you can apply two (or more) competing defeasible inference rules then apply the one that is more specific to the present situation.
I solve the problem of defeaters that are not refuters by modifying Brandom’s account so that we interpret which inferential relations a reasoner implicitly accepts on the basis of how that reasoner would use different expressions in some privileged reasoning context, rather than how she actually uses those expressions in whatever reasoning context she happens to be in. This allows us to interpret a reasoner as implicitly accepting a defeasible inferential relation even if her actual reasoning context is such that she takes the inferential relation to be defeated and therefore does not (or is not disposed to) use the related expressions in the relevant ways.

The end result of my modifications to Brandom’s account is a new way of understanding the nature and role of defeasible inference within a broader theory of reasoning that is social and holistic.
CHAPTER I

General Introduction to Brandom’s Account of Language and Reasoning and his Commitment to its being Predominantly Defeasible

PART 1: An Outline of how Brandom Grounds Meaning in Practice

In this part of the Chapter I will outline Brandom’s general account of how meaning is grounded in practice and present the meta-vocabulary he uses to explain the rest of his overall position. Some of the finer details of the general account presented here will be further explored in Chapters II and III.4

Linguistic Practice and Material Inference

For Brandom a practice’s being inferential is both necessary and sufficient for its being linguistic/conceptual. In fact the meaning (propositional content) of an expression, according to Brandom, is its inferential role. What distinguishes the English speaker’s saying ‘That is red’ from the parrot’s saying it is the English speaker’s also being able to draw various inferences from what she says, such as “That is coloured”. In this way the English speaker grasps the meaning (the inferential role) of the expression, while the parrot does not. Similarly, an iron bar may reliably distinguish those environments that

4 My discussion of Brandom’s theory of language and reasoning is directed at the position that Brandom puts forward in his books Making It Explicit and Articulating Reasons. I will also draw from Brandom’s book Between Saying and Doing, which undertakes a slightly different project, but still provides useful insights into and, at times, elaboration upon Brandom’s position in Making It Explicit and Articulating Reasons. For brevity, I will refer to these books as ‘MIE’, ‘AR’, and ‘BSD’.
are wet from those that are not by rusting or not rusting, but it does not have the concept ‘wet’ because it cannot draw the appropriate conclusions that follow from it.\(^5\) A practice may therefore be reliable and quite complex, but if it is not interpretable as inferential then according to Brandom it is not linguistic (though like Wittgenstein’s block game it could still be verbal). “What makes a classification deserve to be called conceptual classification is its inferential role.” (MIE 89)

The sort of practices that Brandom describes as inferential are not limited to those involving logically valid inferences. Consider the following examples:

(1) \textit{The ball is entirely red}  
\textit{The ball is not green}

(2) \textit{Fido is a dog}  
\textit{Fido is a vertebrate}

(3) \textit{Thunder now}  
\textit{Lightning earlier}

For the above three inferences to be logically valid they would each respectively need the following additional premises:

(a) \textit{IF the ball is entirely red THEN the ball is not green}  
(b) \textit{IF Fido is a dog THEN Fido is a vertebrate}  
(c) \textit{IF Thunder now THEN Lightning earlier}

The sort of inference that is central to Brandom’s account is \textit{material} inference, such that the three arguments above have a sort of goodness just as they stand without adding conditionals (a),(b), or (c) as premises. Inferences can therefore be materially good without being formally valid. (MIE 97-98)

\(^5\) Something need not be able to make all or even most of the correct inferences a concept stands in, but they must be able to make many of them. Understanding a concept is not an all or nothing affair.
Because the meaning (or conceptual content) of a sentence is its material inferential role, one need not be able to use conditionals or other logical vocabulary in order to be a language user. The function of logical vocabulary is to make explicit some of the material inferential roles that we already implicitly associate with different sentences. Using again the examples above, the conditionals (a),(b), and (c) make explicit the material inferential relations implicit in (1), (2), and (3) respectively. The benefit of being able to explicitly formulate material inferential relations in this way is that it allows language users to reflectively talk about and scrutinize what they were doing implicitly and unreflectively. Brandom writes,

The material inferences codified in subjunctive conditionals are inferential involvements that are essential to the contents of the concepts used in science and everyday life. These are not logically valid inferences. But logical vocabulary, subjunctive conditionals, can be used to express these material inferential relations. Without such vocabulary, the inferences can still be endorsed. With it those content generating inferential endorsements can be made explicit as the content of a claim or propositional endorsement. (MIE 103-104).

We can make explicit the meaning of a sentence by expressing the premises leading to that sentence and the conclusions leading from it. These correspond to the sentence’s appropriate circumstances and consequences of application. When these have been made explicit we can evaluate the appropriateness of the concept itself on the basis of whether we do in fact think that it is appropriate to make inferences from the
circumstances of application to the consequences of application. This does not consist of assessing whether or not the concept is used correctly but of assessing whether or not the concept is a good one—one that should be part of our repertoire of concepts. Brandom uses ‘boche’ as an example of a concept that incorporates materially bad inferences. Its circumstances of appropriate application (such as in sentences ‘He is boche’) are when it is applied to any male person of German descent and its consequence of application is that the person is prone to brutality and violence. But because Germans are not in fact prone to violence and brutality, the inferences involved in ‘boche’ are materially bad and therefore the concept itself should be rejected (not used). (MIE 125-130). Until we can make material inferences explicit we cannot critically discuss and evaluate the correctness of the concepts we use.

**Assertion, Commitment and Entitlement**

When Brandom says that a practice is linguistic if and only if it is inferential, he tells us very little unless and until he further explains what structure a practice must have for it to be inferential. It is for this reason that Brandom introduces what he calls the game of giving and asking for reasons (what I will often refer to simply as the ‘reasons game’).

What makes something that is done according to a practice —for instance the production of a performance or the acquisition of a status— deserve to count as *inferring*? The answer developed here is that inferring is to be distinguished as a certain kind of move in the game of giving and asking for reasons. (MIE 158)
Sentences have the material inferential relations that they do as a result of their having the role that they do in the game of giving and asking for reasons. Sentences have a particular role in the reasons game in virtue of how participants *keep score* on moves involving those sentences.

Assertion is the fundamental type of move in the reasons game because only assertions can stand in need of reasons (justification), and can provide reasons for other assertions. (MIE 167) When someone makes an assertion in the reasons game they undertake a commitment and claim\(^6\) entitlement to that commitment. Commitments and entitlements are deontic (and therefore normative) statuses which people have by being appropriately attributed as having.\(^7\) (MIE 162)

For the fundamental concept of the metalanguage employed in specifying the model of assertional practice is that of the deontic attitude of attributing a commitment. For the deontic attitude of undertaking a commitment is definable in terms of attribution: undertaking a commitment is doing something that licenses or entitles others to attribute it. (MIE 196)

Someone’s asserting one sentence rather than another often makes it appropriate to attribute one commitment rather than another. One can also assert a variety of different

---

\(^6\) Brandom frequently talks about commitments that are self-attributed and calls these acknowledged commitments. He rarely talks about entitlements that are self-attributed, but when he does he refers to them as the entitlements the person *claims*. (MIE 185-186) Unfortunately, ‘claims’ may not be the best word because it already plays so many easily confusable roles. When I say that someone S *claims* entitlement to some sentence P, I am not saying that S undertakes commitment to the sentence “S is entitled to P”. All I am saying is that when S asserts P, she implicitly takes herself to be entitled to P.

\(^7\) This goes for commitments generally and not just to assertional commitments.
sentences to undertake one particular commitment (for example, “Bill ate the cake”, and “The cake was eaten by Bill”).

A terminological point: Following Brandom I will often use the expression ‘claim’ to refer to a potential assertional commitment (roughly a propositional content) that various sentences may express. When I talk about claim \( p \), I am not talking about any particular act of claiming or asserting \( p \) or anyone’s actual commitment to \( p \), but rather about what is potentially claimed or asserted, or committed to. It is a possible move in the reasons game that someone could make. Sometimes I will rather loosely say things like “John is committed to the claim ‘Snow is white’”. This should be understood as: John has undertaken the commitment that is expressed by the sentence “Snow is white”.

An assertion involves a commitment in the sense that by making it, one is responsible (or obliged) to do certain other things. First, one is obliged to acknowledge any further commitments entailed by the commitment undertaken in that assertion. Second, one is responsible to demonstrate how she is entitled to that commitment when such a demonstration is appropriately demanded. For example, if S1 asserts “The cat is on the mat” S1 undertakes a commitment to that claim to the effect that she is obliged to

---

8 Like meanings, commitments are individuated by inferential role, which will then be cashed out in terms of scorekeeping significance. Brandom does not hold that assertions are commitments to the asserted sentence itself. Rather the asserted sentence expresses (from a certain perspective) what exactly is being committed to in the assertion. Otherwise we lose any notion of sameness of meaning that goes beyond sameness of sentence. That someone takes assertions of a particular sentence to express a particular commitment is a sort of expressive commitment—something that I will not be addressing here.

9 Brandom does not hold the position that persons can grasp propositional contents independently of being able to use sentences that express them. If two different sentences have the same content, it is not because the two sentences both stand in a certain relation to some third ‘content’ entity. Rather they have the same content because they both play the same role in the reasons game (and therefore have the same inferential role).
acknowledge any commitments that follow from it, such as “There is a mammal on the mat” and “The cat is on something”. If S1’s entitlement to undertake her commitment is appropriately challenged, then S1 is also obliged to demonstrate her entitlement either by making further assertions that act as premises leading to her commitment as a conclusion or by deferring to someone else’s authority. In our example this means that S1 could demonstrate her entitlement either by asserting “I hear the cat meowing at the door and there is a mat at the door”, or by asserting, “S2 told me so”. Being entitled to a claim allows one (and one’s hearers) to use that claim as a premise from which to make various inferences. What is crucial here is that the significance of being committed or entitled to one claim in the reasons game is understood in terms of being committed or entitled to other claims. Brandom writes:

What is done in asserting—the pragmatic significance or effect of producing an assertional performance—consists in the way in which, by authorizing particular further inferentially related performances and undertaking responsibility to produce yet other inferentially related performances, asserters alter the score interlocutors keep of the deontic statuses (commitments and entitlements) of their fellow practitioners.

(MIE p.173)

It is of course possible for persons to undertake commitments to which they are not in fact entitled, and to take themselves to be entitled to claims that they do not in fact acknowledge commitment to. The former involves persons undertaking commitments inappropriately (or more loosely, holding beliefs that require but lack justification); the
latter involves persons thinking that they have reasons that make it permissible (but carry no obligation) to undertake certain commitments.

**Scorekeeping in the Reasons Game**

Language users keep score in the reasons game by keeping track of each other’s commitments and entitlements to claims. This ‘keeping track’ consists of actively attributing commitments and entitlements to oneself and one’s interlocutors.

The inferential relations that a scorekeeper associates with a claim just consist of the ways that that scorekeeper takes commitments and entitlements to that claim to impact commitments and entitlements to other assertions in that context. This is not to say that scorekeepers in the reasons game have an *explicit* set of rules on how to keep score when different assertions are made. Scorekeeping is primarily done implicitly and unreflectively. A scorekeeper can implicitly endorse (undertake commitment to) a claim *p*’s inferentially relating to a claim *q* just by being disposed to add or remove commitments or entitlements to *q* when a commitment is undertaken to *p*. (Which particular type of inferential relation someone counts as endorsing as a result of their having a particular scorekeeping disposition is something that I will discuss later in Chapter III). Brandom writes:

> For at any stage, what one is permitted or obliged to do depends on the score, as do the consequences that doing has for the score. Being rational—understanding, knowing how in the sense of being able to play in the game of giving and asking for reasons—is mastering in practice the
evolution of score. Talking and thinking is keeping score in this sort of
game. (MIE p.183)

Elsewhere Brandom also writes,

Inferential connections among claims are understood in turn
pragmatically, in terms of *consequential* relations among the attitudes by
means of which score is kept on commitments and entitlements to
commitments—how attributing one commitment entails attributing others,
precludes entitlement to others, and so on. (MIE p.472 emphasis in
original)

**PART 2: Setting the Stage**

In the remainder of this Chapter I will be setting the stage for the two arguments
that I will present against Brandom’s account in the next two chapters.

**The General Strategy:**

Brandom introduces the reasons game in order to help explain what makes a
practice count as inferential and to show how certain sorts of practices can institute
material inferential relations. In order to avoid giving a circular explanation, Brandom
must not appeal to inferential relations when he articulates the structure of the reasons
game. The vocabulary used to characterize the reasons game (primarily involving
normative attitudes—attributions of deonic statuses) therefore acts as a sort of
metavocabulary that is used to (among other things) articulate the circumstances and
consequences of someone’s having implicit commitments to Brandom’s three types of
material inferential relations: commitment-preserving relations, entitlement-preserving
relations, and incompatibility relations.\footnote{Brandom also appeals to reliability inferences when discussing perception (language-
entry moves) and action (or language-exit moves). I will be focussing solely upon
theoretical reasoning (intra-linguistic moves) which does not make use of reliability
inferences.}

Once we recognize that the scorekeeping vocabulary is a sort of metavocabulary
that can be used to present the circumstances for and consequences of attributing
inferential commitments, we can distinguish (1) Brandom’s commitment to use the
scorekeeping metavocabulary, from (2) Brandom’s commitment to a particular
characterization (of the circumstances and consequences), in terms of that vocabulary.
The arguments against Brandom’s account that I will be presenting in the next two
chapters are not against the former, but only against the latter. In fact, I will be using
Brandom’s own scorekeeping vocabulary to show that Brandom’s three types of material
inferential relations themselves incorporate materially bad inferences. Here is one
strategy for doing this:

1) Describe an example of good reasoning using the scorekeeping vocabulary
(purely in terms of how a scorekeeper should attribute and withdraw attributions
of commitments and entitlements at various points in the example).

2) Identify the inferential commitments the scorekeeper in the example would
have according to Brandom’s account of the appropriate circumstances for
attributing inferential commitments.

3) In light of the inferential commitments identified in Step 2, identify how the
scorekeeper in the example would be obliged to update score according to
Brandom’s account of the consequences of someone’s being committed to the different types of inferential relations.

4) If in step 3 the scorekeeper’s inferential commitments oblige her to update score in a way that conflicts with how the scorekeeper ought to keep score (as described in Step 1) then Brandom’s three types of material inferential relations incorporate materially bad inferences.

In the next two chapters I will use this strategy to show that Brandom’s three types of inferential relations incorporate materially bad inferential relations when applied to Brandom’s own example of defeasible reasoning. For the remainder of this chapter I will explain what defeasible reasoning is, show that Brandom is committed to reasoning being predominantly defeasible, and present Brandom’s most commonly used example of good defeasible reasoning—an example which will be used throughout the remainder of the dissertation.

**Defeasible Reasoning and its Predominance**

In addition to not being formally valid, many of the inferences made in the reasons game are defeasible and nonmonotonic. Reasoning that is monotonic, like the sort seen in traditional deductive logic, requires that whatever can be inferred from one set of premises can necessarily be inferred from any superset of it. For example, in classical logic we can infer $q$ from the set of premises $\{p, p \rightarrow q\}$, and we will continue to be able to do so regardless of what additional premises might be added. Reasoning where this does not hold is nonmonotonic.
While many philosophers are not familiar with nonmonotonic reasoning, there is a significant amount of research in AI, cognitive science, and formal epistemology about its importance and how we should formalize it. For example, John McCarthy, one of the early AI theorists who first identified the frame problem, argues that “humans use such ‘non-monotonic’ reasoning and that it is required for intelligent behaviour”. (145) So much of our reasoning (both practical and theoretical) depends upon our being allowed to ignore or make assumptions about whether or not certain conditions hold, unless and until we have positive reasons for doing otherwise. Brandom writes

material inference is not in general monotonic—even on the theoretical side. It can be in special cases, say, in mathematics and fundamental physics. But it never is in ordinary reasoning and almost never in the special sciences. (Reasoning in clinical medicine, for instance, is resolutely nonmonotonic.)

Consider the arguments that are codified in the following conditionals:

1. If I strike this dry, well-made match, then it will light. \((p \rightarrow q)\)
2. If \(p\) and the match is in a very strong electromagnetic field, then it will not light. \((p \& r \rightarrow \neg q)\)
3. If \(p\) and \(r\) and the match is in a Faraday cage, then it will light. \((p \& r \& s \rightarrow q)\)
4. If \(p\) and \(r\) and \(s\) and the room is evacuated of oxygen, then it will not light. \((p \& r \& s \& t \rightarrow \neg q)\).
The reasoning we actually engage in always permits the construction of inferential hierarchies with oscillating conclusions like this. (AR 87-88)

Here we not only see Brandom claiming that nonmonotonic reasoning is the most common sort of reasoning, we also see Brandom’s most frequently used example of nonmonotonic (or defeasible) reasoning. Both of these will play central roles in the arguments being put forth in this dissertation.

It is crucial to see that nonmonotonicity (in its strictest sense) is not the only important and interesting feature of the reasoning in Brandom’s match-lighting scenario. The scenario doesn’t just show us that conclusion \( q \) inferentially follows from premise \( p \), but not from the larger set of premises \( \{p, r\} \), which is a violation of monotonicity. It also shows us that \( q \) will inferentially follow from a third even larger set of premises \( \{p, r, s\} \) that contains the two smaller sets. It shows that as we add new claims to our set of premises we may not be permitted to infer what we could previously infer from the smaller set, AND that as we add even further claims to our new set of premises we may reacquire permission to infer what we previously were not permitted to infer. I will call this latter feature entitlement recovery.

Clearly, any instances or patterns of reasoning that allow entitlement-recovery will also be nonmonotonic, but it is not the case that every instance or pattern of reasoning that is nonmonotonic will also allow entitlement-recovery. An account of reasoning would be nonmonotonic if it was able to describe reasoners as appropriately inferring some claim \( p \) from a set of premises \( A \), but not from its superset \( A' \). But it is
possible that this account of reasoning would not have any means to describe reasoners as appropriately inferring \( p \) from any superset of \( A' \).

While *defeasible* reasoning is often taken to be one and the same as *nonmonotonic* reasoning, I take it that what Brandom commits himself to in the quotation provided above is more than just the predominance of nonmonotonic reasoning in the strictest sense. He is committing himself to a sort of defeasible reasoning that is nonmonotonic and allows entitlement-recovery. This is made especially apparent in the last sentence in the text quoted from Brandom above, “The reasoning we actually engage in always permits the construction of inferential hierarchies with oscillating conclusions like this” (AR 88).

An inferential description of some instance of nonmonotonic reasoning would typically employ some *defeasible* inferential relation and a *defeating* inferential relation for the defeasible relation. An inferential relation is defeasible if it has defeating conditions, which if satisfied, would prohibit someone from inferring the inferential relation’s conclusion from its premises. One can conveniently think of these defeating conditions as ceteris paribus (all other things being equal) clauses. For example, we may say that *if the match is struck and all other things are equal then the match will light*. If other things are not equal, perhaps because the match is in a strong electromagnetic field, then we may not infer ‘the match will light’ from the premise ‘the match was struck’, but the inferential relation itself remains good.

A claim is a defeater for an inferential relation (thereby being one of the claims satisfying that inferential relation’s defeating condition) only if that claim stands in certain inferential relations. If \( p \) is a defeater for defeasible inferential relation \( IRa \), then
there must be some inferential relation \( IRb \) that is responsible for \( p \)'s satisfying \( IRa \)'s defeating condition. For example, there must be some inferential relation that makes ‘Tweety is a penguin’ a defeater for the inferential relation leading from ‘Tweety is a bird’ to ‘Tweety can fly’ (such as an inferential relation leading from ‘Tweety is a penguin’ to ‘Tweety cannot fly’). We will call inferential relations that are responsible for a claim’s satisfying a defeat condition \textit{defeating} inferential relations. While defeating inferential relations are responsible for other claims’ being defeaters, they are not themselves defeaters. Therefore, in the present example we would say that the inferential relation from ‘Tweety is a penguin’ to ‘Tweety cannot fly’ is a \textit{defeating inferential relation} (but is not itself a \textit{defeater}) for the inference from ‘Tweety is a bird’ to ‘Tweety can fly’.

For those unfamiliar with defeasible reasoning there is often the temptation to try to turn defeasible inferences into indefeasible ones. For example, one may be tempted to say that the inference from ‘The match is struck’ to ‘The match will light’ is simply a bad inference unless we rule out all those conditions where the match will not light when struck, for example,

\[(\text{Match is struck} \& \text{Match is not in a strong EM field} \& \text{There is oxygen in the room}) \rightarrow \text{The match will light}\]

The main reason why many theorists interested in defeasible reasoning will treat an inference from ‘Match is struck’ to ‘Match will light’ as being a good one without such supplementation is that reasoners often have to make decisions or form beliefs based
on limited knowledge. It is too much to ask of reasoners that they rule out all the possible defeating conditions in which striking a match will not result in its lighting (such as the match’s being in a strong EM field), and, in addition, rule out all the conditions in which other conditions (such as the match’s being in a Faraday cage) defeat those defeating relations. Brandom himself, largely because of his holistic commitments (which will be discussed at length in Chapter III), claims that we cannot even in principle identify all the defeating conditions for a defeasible inferential relation.

The problem is not just that we would need an infinite list of conditions being ruled out—though that is true. It is that the membership of such a list would be indefinite: we do not know how to specify in advance what belongs on the list.(AR 88)

Two Different Senses of ‘Defeasible’

One of the problems that arises when we try to discuss defeasible reasoning is that ‘defeasible’, ‘defeat’, and expressions related to them are often used in relevantly different ways. Suppose we say that someone S’s reason for $q$ was defeated. We might mean:

-S no longer accepts whatever beliefs or rational relations were necessary for her being previously justified in believing $q$.

-S continues to accept whatever beliefs and rational relations formerly justified her belief $q$, but such reasons were inconclusive, and now S has an even stronger reason against $q$. 

Corresponding to these two different senses of ‘defeat’ are two different ways that one can understand reasoning as being ‘defeasible’. Defeasibility might consist of either

1) *The acceptance or non-acceptance of* particular beliefs and the rational relations between those beliefs

Or

2) *Some property belonging to the rational relations* between beliefs.

If it is the *acceptance* of particular beliefs or rational relations that is said to be defeasible, then defeasible reasoning largely amounts to reasoning that allows *revision* (or ‘contraction’—removing beliefs from one’s set of beliefs). I will call this the *refutable* sense of ‘defeasible’.

Alternatively, ‘defeasibility’ can be understood to be *a property belonging to some relations of rational support* between sets of premises and conclusions. On this perspective *p* can be a defeasible reason for *q*, by being a (deductively) inconclusive reason for *q*—allowing that there may be other (potentially stronger) reasons against *q*. I will call this the *inconclusive* sense of ‘defeasibility’. By acknowledging that the relation of support leading from *p* to *q* is defeasible we are not rationally obliged to cease accepting this relation in contexts where we accept *p* but not *q* (perhaps because of some stronger reason against *q*).

**The Defeat/Refute Distinction**

*The key critical claim that I am making in this dissertation is that we cannot use Brandom’s three types of material inferential relations to adequately describe someone*
as correctly engaging in reasoning that is defeasible in the inconclusive sense.

Therefore, unless I say otherwise, I will for the remainder of this dissertation only use ‘defeasible’ (and its associated expressions ‘defeat’, ‘defeater’, etc.) in the inconclusive sense.

Because defeasibility will be treated as a property of certain types of inferential relations (rather than as a property of one’s acceptance of propositions or inferential relations), it will only make sense to say of inferential relations (rather than of propositions) that they can be defeated. One might withdraw (or be rationally obliged to withdraw) one’s commitment to a proposition as a result of the inferential relation justifying her commitment becoming defeated, but the proposition itself is not for this person defeated, but refuted. We can roughly understand a refuter for some inferential relation or claim as being a sort of un-entitler for that claim or inferential relation. Likewise we can understand some claim p’s being refuted for someone S as S’s being un-entitled to p. 11 The refutability of assertional commitments and the inferential relations between them is important to the present topic (and to Brandom’s own project), but it only confuses matters to say that a particular proposition or inferential relation is defeated when one ceases accepting it.

One of the key differences between an inferential relation’s being defeated and its being refuted is that the inferential relation can be defeated for one person but not another, while the inferential relation’s being refuted for one person necessarily means it is refuted for everyone. For example, the inference from ‘The match is struck’ to ‘The match will light’ might be defeated for someone committed to the claim ‘The match is in

---

11 In part 1 of Chapter V, I explain in much more detail the relationships between reasons for, reasons against, defeaters, refuters and the like.
a strong EM field’, but not defeated for someone who isn’t committed to this claim. Even for someone committed to the claim ‘The match is in a Strong EM Field’ there is an important sense in which ‘The match is struck’ remains a good (but defeated) reason for accepting ‘The match will light’. This is not how it works for inferential relations that are refuted. If an inferential relation is refuted then it is materially bad. And just as propositions cannot be false for one person but not another, neither can inferential relations be materially bad for one person but not another. For example, the inferential relation from ‘Anscombe is a woman’ to ‘Anscombe does philosophy poorly’ is materially bad for everyone regardless of what other claims they are committed to (and regardless of whether or not they recognize the inferential relation as being materially bad).

In light of this discussion of the distinction between an inferential relation’s being refuted and defeated, we can draw a corresponding distinction between an inferential relation’s refuters and defeaters:

If some claim \( p \) is a refuter for some inferential relation \( IRx \), then it follows that if \( p \) is true then \( IRx \) is materially bad for everyone regardless of whether or not they are committed to \( p \) (allowing that the inferential relation that makes \( r \) a refuter could itself be defeated).

If some claim \( p \) is a defeater for some inferential relation \( IRx \), then it follows that if someone S is committed to \( p \) then \( IRx \) is defeated for S (recognizing of course
that the inferential relation that makes \( r \) a defeater could itself be defeated). Even if \( p \) is true, it is possible for \( IRx \) to be materially good.

Because inferential relations can remain materially good (hence unrefuted) even when defeated, it follows that the set of claims that are defeaters for an inferential relation will not be a subset of the set of claims that are refuters for an inferential relation. That is, every defeasible inferential relation must have possible defeaters which are not refuters. This will be one of the central premises in the argument given in Chapter III.

At this point, one may quite appropriately ask how or why an inferential relation could remain materially good (for everyone) even when one of its defeaters is true. Answering this question in the context of Brandom’s broader account of language and reasoning is not at all easy and is one of the central aims of this dissertation. A common way to try to answer this question is to treat defeasible inferential relations as rough (rather than strict) generalizations about classes of things, events, or contexts. For example, even if a match is in a strong EM field, we might continue to accept the inference from ‘The match is struck’ to ‘The match will light’ because most matches will light when struck, or generally matches light when struck, or even this particular match \( \text{would in most contexts} \) light when struck. As we will see later (in Chapter IV), however, there are features of Brandom’s account that prevent us from taking these more familiar routes. For present purposes we thankfully don’t need to know how or why an inferential relation can be materially good even if one of its defeaters is true. We need only recognize that a consequence of this is that defeasible inferential relations must have defeaters which are not also refuters.
To see that the match-lighting scenario is an example of reasoning that is
defeasible (rather than just refutable) we need only recognize that some of the inferential
relations in it have defeaters that are not also refuters. Even if the match is in a Faraday
cage, there remains a very important sense in which those who are unaware of this should
still refrain from concluding ‘The match will light’ from the premises ‘The match is
struck’ and ‘The match is in a strong EM field’. This shows that the claim ‘The match is
in a Faraday cage’ is (in this case) a defeater without also being a refuter.

Preliminary Sketch of the Two Arguments for Why Brandom’s Three types of
Inferential Relations Incorporate Materially Bad Inferences.

Earlier I outlined the general strategy for how we can use Brandom’s
scorekeeping metavocabulary to argue that Brandom’s three types of inferential relations
incorporate materially bad inferences. After that I presented Brandom’s most commonly
used example of good defeasible reasoning, this being the match-lighting scenario. Along
the way I have also identified two features belonging to the match-lighting scenario that
commonly arise in other instances of good defeasible reasoning. The first is entitlement-
recovery, whereby one can gain, lose, and reacquire entitlement to a commitment as
result of one’s consecutively acquiring new commitments. The second feature is that
some of the inferential relations involved in the reasoning have defeaters which are not
refuters (and therefore the match-lighting scenario is an example of reasoning that is
defeasible in the inconclusive sense).

In Chapter II I will present Brandom’s account of the consequences of a
scorekeeper’s being committed to each of his three different types of inferential relations.
I will then use the match-lighting example to show that this account does not allow entitlement-recovery (not without having the scorekeeper take some of the inferential relations to be refuted instead of defeated). This argument is largely independent of Brandom’s account of the appropriate circumstances for attributing inferential commitments, because it shows that for any combination of inferential relations, a scorekeeper who is committed to them would be obliged to update score incorrectly (unless the scorekeeper were at some point to take some of the inferential commitments to be refuted rather than defeated). I will call this the problem of entitlement-recovery.

In Chapter III I will present Brandom’s account of the appropriate circumstances for attributing inferential commitments to a scorekeeper. I will then use the match-lighting example to show that Brandom’s account of the circumstances does not allow us to interpret a scorekeeper as taking a claim to be a defeater for an inferential relation without also having us interpret that scorekeeper as taking the claim to be a refuter for the same inferential relation. I will call this the problem of defeaters that are not refuters.

**PART 3: The Gap in the Secondary Literature**

**Common Objections to Brandom’s Account in the Secondary Literature**

Given the scope and originality of Brandom’s overall account of language and reasoning, it is not surprising that Brandom’s account has been criticized on such a wide range of issues. While the arguments raised against Brandom’s account are diverse, certain features of the account tend to be criticized much more than others.

One of the most common worries that people have with Brandom’s account is that it fails to explain (or perhaps even allow) for reasoning’s being objective. According to
the typical representationalist view of language and reasoning, a belief or proposition is objectively true if and only if it successfully represents some mind independent state of affairs (at least in standard cases where the belief or proposition is not about mental states). Brandom’s account of reasoning and meaning threatens to undermine this understanding of objectivity by explaining reasoning and meaning in terms of something that looks very mind dependent—normative social practices.

Sven Rosenkranz tries to argue that Brandom’s account does not allow for objective truth on grounds that the account does not allow that there could be propositions that are true, but which we do not have warrant to believe. The worry is that Brandom does not allow for truth to be independent of what people believe or are entitled to believe—indeed independent of something mind independent.

Bernd Prien, in his 2010 paper “Robert Brandom on Communication, Reference, and Objectivity”, argues that even though Brandom succeeds at showing that conceptual norms are independent of anyone’s attitudes, Brandom fails to show that conceptual norms are objective. He writes, “Objectivity of conceptual norms, by contrast, amounts to the positive claim that it is the world that determines deontic statuses.” (Prien 454, italics added) This world dependence, Prien argues, is what Brandom’s account in MIE fails to

---

12 See “Farewell to Objectivity: A Critique of Brandom” (2001)
show. John Haugeland provides a similar, but very brief, argument against Brandom’s account in his book “Having Thought”.

Another common objection raised against Brandom is that his account of meaning does not allow for successful communication to be possible. This is one of the central objections raised by Fodor and Lepore. As we shall see in Chapter III, Brandom accepts a sort of meaning holism such that a reasoner’s inferential commitments depend (in an indeterminate way) on that reasoner’s assertional commitments. In short, what someone means will depend upon that person believes.

Fodor and Lepore (2007) argue that because different speakers will have different beliefs, it follows from Brandom’s account that one and the same sentence will mean different things to different speakers. This problem could be avoided, they argue, if Brandom’s account allowed a distinction between those inferential relations that are meaning-constitutive and those that are not (a sort of analytic-synthetic distinction). But because there is no such analytic-synthetic distinction, argue Fodor and Lepore, there will be no identity of meaning between different speakers, and therefore communication becomes impossible.

Daniel Whiting in his 2008 paper “Meaning Holism and De Re Ascription” argues similarly to (though much more carefully than) Fodor and Lepore. Whiting points out that even though Brandom rejects the analytic-synthetic distinction, Brandom tries to

---

13 Prien goes on to argue that aspects of Brandom’s account in BSD can solve this problem.
14 See chapter 13, “Truth and Rule-Following”. Haugeland’s argument is largely directed at social pragmatism generally, but he does address features of Brandom’s account in particular 357-358 (end note 14).
15 See Fodor and Lepore 2007.
explain successful communication with his account of de re ascriptions (something Fodor and Lepore largely ignore). The problem, Whiting argues, is that Brandom’s account of de re ascription does not establish how different speakers can have a shared perspective on meaning, but only shows how each scorekeeper can make sense of other people’s utterances relative to his or her own sets of beliefs—relative to one unshared perspective on the utterance’s meaning. Prien (2010), mentioned above, discusses many of the same issues as Whiting’s, but does not find this to be a problem for Brandom.

Another objection raised by Fodor and Lepore against Brandom’s account focuses on the compositionality of language. According to Fodor and Lepore (2001), an account of meaning can only be successful if it explains the compositionality of language, whereby the meaning of a whole sentence is determined by (and can be explained in terms of) the meanings of its parts. Fodor and Lepore argue that because the inferential roles of whole sentences (and hence meanings on Brandom’s account) are not determined by the inferential roles of sentence parts, inferentialist theories of meaning like Brandom’s do not allow for (nor explain) the compositionality of language.

In Response to Fodor and Lepore’s arguments from compositionality, Mark McCullagh in his 2003 paper “Do Inferential Roles Compose?” argues that Brandom does in fact provide an account of how subsentential expressions have inferential roles and how these can in fact determine the inferential roles of sentences.

The foregoing was just a brief sample of the sorts of arguments commonly raised against Brandom’s account of language and reasoning. It is well beyond the scope of this dissertation to assess whether or not these arguments are successful, or what sort of
impact they would have if they were successful.\textsuperscript{16} I will examine Brandom’s holism and features of his account of objectivity in Chapter III, but I will not spend much any time examining Brandom’s account of communication, or whether his account allows compositionality.

**What’s Missing From the Secondary Literature**

What is surprisingly absent from the secondary literature is any critical discussion about whether or not Brandom gives an adequate account of defeasible reasoning. I suspect that this may be because most of the people who would be interested in Brandom’s project would not typically be the same people who are interested in defeasible reasoning.

If one were to equate reasoning’s being holistic with reasoning’s being defeasible then one might be led to think that the secondary literature discussing Brandom’s holism constitutes discussion of Brandom’s account of defeasible reasoning. However, this conflates the crucial distinction between reasoning which allows *refutability* and reasoning which involves *inconclusive* relations of rational support. It is the latter that is central to this dissertation.

Even if we look beyond how Brandom’s account handles defeasible reasoning specifically, there has not, as far as I can tell, been any attempt to critically examine the

\textsuperscript{16} For what it is worth, I think that most of the arguments raised against Brandom’s account do not succeed, or if they do succeed they make no real difference to Brandom’s project. Most of the arguments seem to depend on premises which Brandom would not accept, but which are commonly accepted in more standard representationalist accounts (Fodor and Lepore’s arguments are good examples of this).
relations between the circumstances for and consequences of attributing commitment to
the three different types of inferential relations that are core to Brandom’s account.\footnote{John MacFarlane has given some careful attention to the consequences of attributing inferential commitments. His computer program GOGAR (Game Of Giving and Asking for Reasons) simulates how scorekeepers on Brandom’s account are obliged to update score in light of their inferential commitments. See http://johnmacfarlane.net/gogar.html. However, MacFarlane has not, as far as I know, examined the relationship between both the circumstances and consequences of attributing inferential commitments. It is also worth noting that MacFarlane’s program fails to handle Brandom’s match-lighting scenario in the same way that I argue (in chapter II) that Brandom’s account fails.}
CHAPTER II

Brandom’s Account of the Consequences of Attributing Inferential Commitments, and Why it does not Allow Entitlement-Recovery

PART 1: The General Strategy

In this chapter I will present Brandom’s account of the consequences of attributing inferential commitments and then argue that it does not allow entitlement-recovery. My strategy for doing this is the one that I introduced in the previous chapter. I will: (1) identify an example of good defeasible reasoning that involves entitlement-recovery; (2) describe this example using the scorekeeping metavocabulary (purely in terms of a scorekeeper attributing and withdrawing attributions of commitment and entitlement); (3) show that Brandom’s account of the consequences of a scorekeeper’s being committed to each of the three different types of inferential relations would oblige scorekeepers in the example to update score in such a way that it prevents entitlement-recovery.

The example of good defeasible reasoning that I will be using is Brandom’s own match-lighting scenario, which we looked at in the previous chapter. To make it easier to discuss the match-lighting scenario I will for the remainder of this dissertation use the abbreviations below on the left to represent the propositions expressed by the sentences below on the right:

\[ \text{Struck} = \text{The dry, well-made match is struck} \]

\[ \text{Lights} = \text{The match lights} \]
Before describing the example purely in terms of a scorekeeper attributing and withdrawing attributions of commitment and entitlement, I will first present Brandom’s account of the consequences of attributing inferential commitments.

**PART 2: Brandom’s Account of the Consequences of Attributing Inferential Commitments.**

**Introduction to Brandom’s Three Different Types of Inferential Relations**

The consequence of someone’s being committed to an inferential relation is that person’s being obliged to update score (on herself and others) in a particular way in response to certain moves being made in the reasons game. How exactly someone is obliged to update score depends in part upon what type of inferential relation it is that he or she is committed to. The three different types of inferential relations that Brandom introduces and makes use of are: commitment-preserving (committive), entitlement-preserving (permissive), and incompatibility.

A claim $p$ *committively* entails $q$, which I will express as `$p \rightarrow q$’ if commitment to $p$ entails *commitment* to $q$. Scorekeepers committed to $p \rightarrow q$ are therefore *obliged* to attribute commitment $q$ to anyone they attribute commitment $p$ to. Despite these inferential relations being called *commitment*-preserving, commitment to them also has

---

$EMF = $ The match is in a strong electromagnetic field  
$FCage = $ The match is in a Faraday cage  
$Vacuum = $ The room is evacuated of oxygen
an impact on how one is obliged to attribute entitlements. If a scorekeeper S1 is committed to \( p \rightarrow q \) and she attributes to S2 entitlement \( p \) (and not necessarily commitment \( p \)), then S1 is obliged to attribute to S2 entitlement \( q \). Committive inferences are commonly made in deductive reasoning.

A claim \( p \) permissively entails \( q \), which I will express as ‘\( p > q \)’, if commitment and entitlement to \( p \) defeasibly entails entitlement to \( q \). Scorekeepers committed to \( p > q \) are therefore obliged to attribute entitlement \( q \) to anyone they attribute commitment and entitlement \( p \) to (assuming that they don’t also attribute commitments that they take to be incompatible with \( q \)). Permissive inferences are commonly made in inductive reasoning (or any other sort of reasoning that is defeasible).

A claim \( p \) is incompatible with \( q \), which I will express as ‘\( p//q \)’, if commitment to \( p \) precludes entitlement to \( q \) and commitment to \( q \) precludes entitlement to \( p \). Persons committed to \( p//q \) are therefore obliged to withhold attributing entitlement to \( q \) from anyone to whom they attribute commitment to \( p \) (and vice versa). Scorekeepers committed to incompatibility relations are not only obliged to withhold attributing new entitlements, they are also obliged to withdraw previously attributed entitlements if they come to attribute new commitments that are incompatible with them. Incompatibility relations are commonly employed in modal reasoning and in defeasible reasoning (as defeating relations).  

\(^{18}\) Incompatibility relations are not strictly speaking inferential in the sense that one inherits commitment or entitlement to one claim as a result of being committed or entitled to another. They do however express scorekeeping proprieties and for my purposes (and Brandom’s –see MIE p. 189) we can count them as inferential in a broader sense of inferential.
The Scorekeeping Algorithm: A More Detailed Account of the Assertional Consequences of Three Different Types of Inferential Relations and How they Interact

How a scorekeeper’s inferential commitments oblige her to update score in response to a move being made in the reasons game depends in part upon what type of move is being made with that claim—that is, what force is being attached to that propositional content. While assertion is the central type of move in Brandom’s reasons game, one may also disavow commitments, or challenge an interlocutor’s entitlement to some commitment (MIE 191-193). Whether a scorekeeper’s interlocutor asserts $p$ or disavows $p$ will make no difference to the various inferential relations that the scorekeeper takes $p$ to stand in, but it nonetheless makes a crucial difference to how the scorekeeper is obliged to update her interlocutor’s score.

While I acknowledge that a scorekeeper’s inferential commitments will have different scorekeeping consequences depending upon what types of move are being made, I will throughout this dissertation be focusing almost exclusively on the consequences of assertions. Therefore when I say ‘consequences of attributing inferential commitments’ I will mean assertional consequences specifically.

Brandom provides his most detailed account of the assertional consequences of one’s being committed to the three types of inferential relations and how these commitments interact with one another in what I call his scorekeeping algorithm. This is found in MIE chapter three (on pages 190-191) under the heading “How Inferential and Incompatibility Relations among Contents Affect the Score”. I will present the scorekeeping algorithm as a series of numbered steps so that easy reference can be made
to them later (though this is not the way that Brandom presents them). Because this is so central to the argument being developed, I will be quoting extensively.

The context for the algorithm consists of a scorekeeper S1 keeping score on an interlocutor S2 who asserts p.

The way S1’s score [the scores S1 attributes] ought to be transformed is settled by the content of S2’s claim, thought of as its tripartite inferential role in commitment- and entitlement-preserving consequence relations and in incompatibility relations connecting commitments and entitlements.

(MIE 190)

**Step 1: Add commitment p**

To begin with, [S1] must add p to the list of commitments attributed to [S2] (supposing the more interesting case in which [S1] does not already attribute to [S2] a commitment with that content). (MIE 190)

**Step 2: Add commitments to commitment-preserving consequences of p**

[S2] should then add also commitment to any claims q that are committive-inferential consequences of p, in the context of the other claims attributed to [S2]. These will vary depending on the auxiliary hypotheses available, according to what other commitments [S1] already attributes to [S2]. (MIE 190-191)

**Step 3: Remove entitlements that are incompatible with p**
Next the incompatibility relations that $p$ (and so its commitment-inferential consequences) stand in must be consulted to determine which, if any, of the entitlements [S1] previously attributed to [S2] are precluded by the newly attributed commitment. Assertions add new commitments, but they can not only add but also subtract entitlements. (MIE 191)

Step 4: Add entitlements to the conclusions of any undefeated commitment-preserving relations whose premises S2 is already entitled

Then in light of the incompatibility relations associated with all of the commitments attributed to [S2], [S1] can attribute entitlements to any claims that are commitive-inferential consequences of commitments to which [S2] is already taken to be entitled, closing the score under commitment-preserving inferences, where the resulting attributions of entitlement are not defeated by incompatibilities. [Brandom’s endnote: “If one is entitled to $p$ and $p$ commitment-entails $q$, one is entitled to $q$—any entitlement-defeating incompatibilities to $q$ equally defeat entitlement to $p$. (MIE 674 note 43).] (MIE 191)

Step 5: Add entitlements to the conclusions of any undefeated entitlement-preserving relations whose premises S2 is already committed and entitled

Next constrained by the entitlement-precluding incompatibility relations associated with all of the other commitments attributed to [S2], [S1] can attribute entitlements to any claims that are permissive-inferential
consequences of commitments to which [S2] is already taken to be entitled, closing the attributed score under entitlement-preserving inferences not defeated by those incompatibilities. (MIE 191)

Step 6: IF applicable, add entitlement \( p \) to \( S2 \)’s interlocutors

Then \( [S1] \) needs to assess \( [S2] \)’s entitlement to the claim that \( p \), by looking at good inferences having it as a conclusion and premises to which \( [S2] \) is committed and entitled… If \( [S1] \) takes \( [S2] \) to be entitled to \( p \) by [commitment or entitlement preserving relations, by testimony, by default, or by reliability inferences] then \( [S1] \) will take \( [S2] \) to successfully entitle others (including \( [S1] \)) to that claim (in the absence of incompatible defeasors). (MIE 191)

As we reconstruct Brandom’s example of defeasible reasoning in terms of scorekeeping practices, there are two key features of the scorekeeping algorithm that need to be noticed. First, the only type of inferential relations that has defeating conditions (and is therefore defeasible) is entitlement-preserving relations. Recall from Chapter I, that an inferential relation has defeating conditions (and is therefore defeasible) if it has conditions, which, if satisfied, would prohibit someone from inferring that inferential relation’s conclusion from its premises. This is just what we see in Step 5: Even if \( S1 \) were committed to \( p > q \) and attributes commitment and entitlement \( p \) to \( S2 \), there are conditions under which \( S1 \) ought not to attribute entitlement \( q \) to \( S2 \). In this case the
defeating conditions would be satisfied if S1 attributes to S2 a commitment to any claim that S1 takes to be incompatible with \( q \).

This brings us to the second feature of the scorekeeping algorithm that I want to draw attention to: only incompatibility relations can be *defeating* inferential relations. Recall from Chapter I, that *defeating* inferential relations are responsible for claims being defeaters for others inferential relations. They are what make commitment to some claim satisfy the defeating conditions of an inferential relation. Step 5 tells us that the defeating conditions for an entitlement-preserving relation are satisfied if S1 attributes to S2 a commitment to any claim that S1 takes to be incompatible with \( q \). This can be seen where Brandom presents the qualification “not defeated by those incompatibilities” (MIE 191). S1 would only take some claim \( r \) to be a defeater for \( p > q \) if S1 were committed to \( r // q \), which would make \( r // q \) a defeating inferential relation for \( p > q \).

On first glance it may look as though defeasible inference plays a role in Step 4 of the scorekeeping algorithm since Brandom seems to gives the same qualification that he gives for Step 5 “not defeated by incompatibilities” (MIE 191). If this were correct, then it would appear as though commitment-preserving relations are also defeasible. In considering this suggestion, we need to be mindful of the crucial distinction made in Chapter I between the *refutable* and *inconclusive* senses of defeasibility. Recall that on the inconclusive sense of ‘defeasibility’ (the sense adopted in this dissertation) defeasibility is a property of the rational relations between beliefs. It not a property of the acceptance or non-acceptance of beliefs (which is the refutable sense). The text quoted in Step 4 is an instance where Brandom uses ‘defeat’ in the refutability sense of defeasible, rather than the preferred inconclusive sense being employed here. To see this we need
only notice that (1) Brandom is treating ‘defeasibility’ as a property of the scorekeeper’s attribution rather than as a feature of an inferential relation, and (2) it is not an inferential relation that is defeated, but one’s entitlement to a claim.

**PART 3: Why Brandom’s Account of the Consequences of Attributing Inferential Commitments does not Allow Entitlement-Recovery**

**Describing the Match-Lighting Scenario in Terms of Attributing and Withdrawing Attributions of Commitments and Entitlements**

When Brandom presents his match-lighting example as set of conditionals, he neither indicates what types of material inferential relations each of the conditionals are supposed to express, nor whether the antecedents or consequents of the conditionals consist of commitments or entitlements. Filling in these details requires a bit of interpretative work. At this point in the discussion, I will make no attempt to argue that a particular set of inferential relations actually captures the reasoning in the match-lighting scenario. On the contrary, I will be arguing that we *cannot* use Brandom’s three types of inferential relations (as they presently stand) to capture the match-lighting scenario. At present, I only aim to give a plausible interpretation of how Brandom understands his own match-lighting example in terms of his three types of inferential relations.

I think that it is pretty clear from Brandom’s scorekeeping algorithm and some of his other discussions about the three types of inferential relation, that Brandom intends some of the conditionals in the example to express entitlement-preserving relations and others to express incompatibility relations.
That Brandom likely intends some of the conditionals in the example to express entitlement-preserving relations emerges from the fact that, on Brandom’s account, entitlement-preserving relations are the only type of inferential relation that can be defeated. Furthermore, when Brandom introduces entitlement-preserving relations in MIE chapter 3, he uses the inference from *Struck to Lights* as an example.

Inductive empirical inferences exploit relations of this genus. The premises of these inferences entitle one to commitment to their conclusions (in the absence of countervailing evidence) but do not compel such commitment. For the possibility of entitlement to commitments incompatible with its conclusion is left open. In this way the claim that this is a dry, well-made match can serve as a justification entitling someone to the claim that it will light if struck. But the premise does not commit one to the conclusion, for it is compatible with the premise that the match is at such a low temperature that friction will not succeed in igniting it. (MIE 168-169)

While it is clear that Brandom takes entitlement-preserving relations to play a crucial role in the match-lighting scenario, it is unlikely that Brandom takes them to be the only type of inferential relation involved. This is because someone’s being committed and entitled to the premises of one entitlement-preserving relation could not result in that person’s losing entitlement to the conclusion of another entitlement-preserving relation. One of the interesting and peculiar features of Brandom’s understanding of entitlement is that one can be entitled to two incompatible claims at the same time. (MIE 675 endnote 44) For example, I may be entitled to undertake commitment to either ‘The shopping mall
will be busy today’ or ‘The shopping mall will be quiet today’ in virtue of my having good reasons for both claims, such as ‘Today is the last Saturday before Christmas’ and ‘There will be a blizzard today’. On Brandom’s view, so long as I do not in fact undertake commitment to either ‘The shopping mall will be busy’ or ‘The shopping mall will not be busy’ then I can be entitled to both claims despite their being incompatible with one another. We see then that the conclusions of entitlement-preserving relations consist only of entitlements (not commitments), and becoming entitled to a claim cannot result in one’s losing entitlement to a different claim.

One can, however, lose entitlement to a claim as a result of one’s becoming committed to another claim. As seen in step 3 of the scorekeeping algorithm, this is precisely the work that is done by Brandom’s incompatibility relations. This shouldn’t be surprising given that (as discussed earlier) incompatibility relations are the only type of inferential relations can that can be defeating relations.

I have now shown that we should interpret Brandom as taking some of the conditionals in the match-lighting scenario to express entitlement-preserving relations, and others to express incompatibility relations. To identify which of the conditionals Brandom takes to express incompatibility relations, and which he takes to express entitlement-preserving relations, we can look to Brandom’s writing, “To assert that p is incompatible with q, one asserts the conditional whose antecedent is p and whose consequent is the negation of q” (MIE 115). This strongly suggests that we should interpret the conditionals which do not have negations in the consequent as entitlement-preserving, and the conditionals with negation in the consequent as incompatibility relations. As a result we have the following inferential relations.
Before we can assess whether a scorekeeper committed to these inferential relations would be obliged to update score correctly, we must first identify what correct score updating would consist of in the match-lighting scenario. Like Brandom’s pre-logical language users, we theorists come to the match-lighting scenario with a sense of how score should be updated in it, without needing to appeal to inferential relations.\textsuperscript{19} When we think about scorekeeping in this way, we only concern ourselves with how a scorekeeper should add or remove commitments and entitlements in response to and in relation to various new commitments and entitlements being consecutively added to her interlocutor’s score. Once we have identified this, we will use it as our standard for assessing whether or not a scorekeeper following Brandom’s scorekeeping algorithm would update score correctly.

Correct score updating in the match-lighting example can be expressed in simple score updating terms using a semi-tabular form, where $S_1$ is the scorekeeper keeping score on another person $S_2$. $S_2$’s commitments and entitlements (according to $S_1$) at any given conversational stage are expressed in three columns. The first column called ‘claims’ indicates the claim that $S_1$ takes $S_2$ to be committed or entitled to. The second and third columns under the headings ‘commitment’ and ‘entitlement’ indicate with an

\textsuperscript{19} One may call this sense our “intuitions”, but I won’t because I am very uncomfortable with appeals to intuitions.
‘X’ whether S2 (according to S1) is committed or entitled to the claim in the same row as the X.

Groups of rows will be divided into different conversational stages, where a new conversational stage begins after a move is made in the reasons game and score is updated in response to that move. In each conversational stage I will list the initial score in the conversational stage, the move made, and the score set that results. When an attribution of commitment or entitlement is withdrawn in the Score Update row, I will indicate so by replacing the ‘X’ with ‘--’. We can understand the type of move that S2 makes at each conversation stage as being assertion, whereby S1 takes S2 to be committed and entitled to the asserted claim. For simplicity, I am going to ignore how exactly S2 acquired entitlement to the various claims that she asserts at different stages.
<table>
<thead>
<tr>
<th>STAGE 1</th>
<th></th>
<th>Commitment</th>
<th>Entitlement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Score:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Struck</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Lights</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Move Made:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMF</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Score Update:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lights</td>
<td></td>
<td></td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STAGE 2</th>
<th></th>
<th>Commitment</th>
<th>Entitlement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Score:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Struck</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>EMF</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Move Made:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCage</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Score Update:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lights</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STAGE 3</th>
<th></th>
<th>Commitment</th>
<th>Entitlement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Score:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Struck</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>EMF</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>FCage</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Lights</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Move Made:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacuum</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Score Update:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lights</td>
<td></td>
<td></td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STAGE 4</th>
<th></th>
<th>Commitment</th>
<th>Entitlement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Result:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Struck</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>EMF</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>FCage</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Vacuum</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Lights</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
The scorekeeping in this reconstruction proceeds just as we think it should and expresses how a theorist who interprets speakers in terms of what their commitments and entitlements are (such as Brandom himself) is committed to characterizing the match-lighting scenario. It will therefore act as the standard by which we can assess the success of Brandom’s account (and later my revised account) of the consequences of attributing inferential commitments.

**Applying Brandom’s Account of the Consequences of Attributing Inferential Commitments to the Match-Lighting Scenario**

I will now argue that if we attributed to a scorekeeper commitment to each of the inferential relations that Brandom’s account suggests then according to Brandom’s account of the consequences, the scorekeeper would be obliged to update score incorrectly. My argument will consist of showing that if a scorekeeper updated score in the match-lighting scenario according to Brandom’s scorekeeping algorithm, then she would update score in a way that conflicted with the score updating in Table 1. For each score update that the scorekeeper S1 makes on her interlocutor S2, I will indicate which step in the scorekeeping algorithm and which inferential commitment obliges S1 to update score in the given manner. This will be listed under the column labelled ‘explanation’.
TABLE 2

Updating Score in the Match-lighting Example
According to one’s Inferential Commitments

S1’s Inferential Commitments:

A) \( \neg g_{18} \ldots \neg g_{18} < 2 \neg g_{18} < 0 \neg g_{18} < 3 \neg g_{18} 5 \neg g_{18} 6 \neg g_{18} 3 \neg g_{18} 8 \neg g_{18} 6 \neg g_{18} 5 = \neg h \neg g_{18} < 2 \neg g_{18} < 1 \)

B) \( \{ \neg g_{18} \ldots 5 \neg g_{18} < 2 \neg g_{18} < 0 \neg g_{18} < 3 \neg g_{18} 5 \neg g_{18} 6 \neg g_{18} 3, \neg g_{18} 3 \neg g_{18} 8 = \neg g_{18} 3 \neg g_{18} 6 \neg g_{18} 5 = \neg h \neg g_{18} < 2 \neg g_{18} < 1 \} \)

C) \( \{ \neg g_{18} \ldots 5 \neg g_{18} < 2 \neg g_{18} < 0 \neg g_{18} < 3 \neg g_{18} 5 \neg g_{18} 6 \neg g_{18} 3, \neg g_{18} 3 \neg g_{18} 8 = \neg g_{18} 3 \neg g_{18} 6, \neg g_{18} 8 \neg g_{18} 5 = \neg g_{18} 5 \} \)

D) \( \{ \neg g_{18} \ldots 8 \neg g_{18} 3 \neg g_{18} 5 \neg g_{18} 6 \neg g_{18} 3 \neg g_{18} 3 \neg g_{18} 6 \neg g_{18} 3 \neg g_{18} 5 \} \)

<table>
<thead>
<tr>
<th>S2’s Score (According to S1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Claims</strong></td>
</tr>
<tr>
<td><strong>Commitment</strong></td>
</tr>
<tr>
<td><strong>STAGE 1</strong></td>
</tr>
<tr>
<td>Initial Score:</td>
</tr>
<tr>
<td>Struck</td>
</tr>
<tr>
<td>Lights</td>
</tr>
<tr>
<td>Move Made:</td>
</tr>
<tr>
<td>EMF</td>
</tr>
<tr>
<td>Score Updated:</td>
</tr>
</tbody>
</table>
| Lights | -- | (3) \( \neg h \neg g_{18} < 2 \neg g_{18} < 1 \)
|**STAGE 2** |
| Initial Score: |
| Struck | X | X | |
| EMF | X | X | |
| Move Made: |
| FCage | X | X | |
| Score Updated: |
| NO CHANGE! | (5) \( \neg h \neg g_{18} < 2 \neg g_{18} < 1 \)
|**STAGE 3** |
| Initial Score: |
| Struck | X | X | |
| EMF | X | X | |
| FCage | X | X | |
| Move Made: |
| Vacuum | X | X | |
| Score Updated: |
| No Change | Not entitled to \( \neg h \neg g_{18} < 2 \neg g_{18} < 1 \)
|**STAGE 4** |
| Overall Result: |
| Struck | X | X | |
| EMF | X | X | |
| FCage | X | X | |
| Vacuum | X | X | |
At Stage 1 the scenario proceeds as we expect. But already at Stage 2 we see that one of the consequences of S1’s being committed to the inferential relations that make up Brandom’s match-lighting example is that she is obliged to withhold attributing entitlement to \( \text{Lights} \) so long as she attributes commitments that are incompatible with it (as required by scorekeeping algorithm step 5). That is, because S1 is committed to \( \{ \text{Struck}, \text{EMF} \} \) \( \triangleleft \) \( \text{Lights} \) and attributes to S2 commitment and entitlement \( \text{Struck} \) and \( \text{EMF} \), S1 is obliged to refrain from attributing entitlement \( \text{Lights} \) to S2 no matter what further commitments and entitlements S1 might take S2 to have. This shows that S1’s inferential commitments oblige her to update score in such a way that makes entitlement-recovery impossible, and therefore they oblige her to update score incorrectly in instances of good reasoning that involve entitlement-recovery (such as the match-lighting scenario).

**Unsuccessful Responses to the Argument that Brandom’s Account does not Allow Entitlement-Recovery**

In an attempt to overcome the problems just raised, one might try to reconstruct the reasoning in the match-lighting example by attributing a different set of inferential commitments to S1. One option would be to take S1 to be committed to inferential relations with smaller premise sets, such as \( \text{EMF} \) \( \triangleleft \) \( \text{Lights} \) instead of \( \{ \text{Struck}, \text{EMF} \} \) \( \triangleleft \) \( \text{Lights} \). This, however, will not be of any help to Brandom. So long as S1 takes S2 to be committed to anything that S1 takes to be incompatible with \( \text{Lights} \) then S1 will be obliged to withhold attributing entitlement \( \text{Lights} \) to S2. And the only way by which
someone can lose entitlement as a result of acquiring a commitment is by means of incompatibility relations. On Brandom’s account, if someone loses entitlement to a claim $q$ because they acquired a new commitment to a claim $r$ then that person cannot reacquire entitlement to the claim $q$ without first withdrawing commitment to the claim $r$ that is incompatible with $q$. This is precisely why Brandom’s account does not allow entitlement-recovery.

One might instead suggest that we attribute to $S1$ commitment to

\{Struck, EMF, $\sim FCage\} // Lights\text{ instead of }\{Struck, EMF\} // Lights\text{. The idea being that the latter of the two inferential relations is only materially good if } FCage \text{ is not true. The problem with this suggestion is that it is simply an attempt to turn defeasible reasoning into indefeasible reasoning. It is an attempt to make inconclusive reasons conclusive. Often it is impossible to do this, both in practice and in principle. As Brandom argued, we cannot in principle specify all and only those conditions in which a defeasible inference will be defeated. Whether or not } EMF \text{ will make a difference to the inference from Struck to Lights will depend upon whether or not } FCage \text{ is present as a premise. But whether or not } FCage \text{ makes this difference, will depend upon whether Oxygen is present as a premise. We have no principled grounds for claiming to have identified all and only those conditions in which a defeasible inferential relation would be defeated because they are always dependent on the presence or absence of further unspecified conditions.}^{20}\n
\hfill

---

20 The inferential holism very coarsely described here will be discussed in much greater detail in Chapter III.
Another way that someone may respond to my argument is by objecting to my interpreting some of the conditionals in the match-lighting scenario as expressing entitlement-preserving relations and others as expressing incompatibility relations. One may argue that Brandom himself uses the same symbol (→) throughout the example and therefore we should treat all the inferential relations in it as entitlement-preserving (since they are the defeasible ones). Instead of treating the conditionals with negated consequents as expressing incompatibility relations, we could treat the negations as part of the content of the consequent. This would give us the following table:
TABLE 3
Updating Score in the Match-lighting Example
According to one's Inferential Commitments (all of which are permissive)

S1’s Inferential Commitments:
A) \( \text{Struck} > \text{Lights} \)
B) \( \{\text{Struck}, \text{EMF}\} > \neg \text{Lights} \)
C) \( \{\text{Struck}, \text{EMF}, \text{FCage}\} > \text{Lights} \)
D) \( \{\text{Struck}, \text{EMD}, \text{FCage}, \text{Vacuum}\} > \neg \text{Lights} \)

<table>
<thead>
<tr>
<th>S2’s Score (According to S1)</th>
<th>Claims</th>
<th>Commitment</th>
<th>Entitlement</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STAGE 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Score:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{Struck} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>( \text{Lights} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Move Made:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{EMF} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Score Updated:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \neg \text{Lights} )</td>
<td>X</td>
<td>X</td>
<td></td>
<td>((5){\text{Struck}, \text{EMF}} &gt; \neg \text{Lights})</td>
</tr>
<tr>
<td><strong>STAGE 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Score:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{Struck} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>( \text{Lights} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{EMF} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>( \neg \text{Lights} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Move Made:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{FCage} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Score Updated:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Change</td>
<td></td>
<td></td>
<td></td>
<td>Already Entitled to ( \neg \text{Lights} )</td>
</tr>
<tr>
<td><strong>STAGE 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Score:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{Struck} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>( \text{Lights} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{EMF} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>( \neg \text{Lights} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{FCage} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Move Made:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{Vacuum} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Score Updated:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Change</td>
<td></td>
<td></td>
<td></td>
<td>Already Entitled to ( \neg \text{Lights} )</td>
</tr>
<tr>
<td><strong>STAGE 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Score:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{Struck} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>( \text{Lights} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{EMF} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>( \neg \text{Lights} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{FCage} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>( \text{Vacuum} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
The problem that arises in Table 3 is not just that S2 cannot recover entitlement to $q$ or $\sim q$, but that S2 cannot even lose entitlement to $q$ or $\sim q$ as a result of acquiring new commitments. This problem would not be resolved by attributing to S1 a further inferential commitment to $q / \sim q$ because, once again, on Brandom’s model it is entirely permissible for someone to be entitled to two or more incompatible claims (such as $q$ and $\sim q$) so long as one is not committed to any of them.

**Summary of the Problem and the Plan on How to Address It**

The key reason why Brandom’s account of the consequences of attributing inferential commitments does not allow entitlement-recovery is this: *It does not have any types of inferential relations that can be both defeating and defeated* (recall that an inferential relation $IRx$ is a defeating inferential relation for some defeasible inferential relation $IRy$ if $IRx$ is largely responsible for some claim’s being a defeater for $IRy$). The only defeating inferential relations on Brandom’s account are incompatibility relations, but these are not themselves defeasible (as seen in Table 2). The only type of defeasible inferential relations on Brandom’s account is entitlement-preserving relations, but this cannot be a defeating relation (as seen in Table 3). As a result, *Brandom’s three types of inferential relations incorporate materially bad inferences, because no matter what combination of inferential relations we interpret a scorekeeper as being committed to, that scorekeeper will be obliged to update score incorrectly in any instances of reasoning where entitlement-recovery should occur.*
If we are to modify Brandom’s account of the consequences of attributing
inferential commitments so that it will allow entitlement-recovery then we will either
have to modify incompatibility relations so that they can be defeated, or modify
entitlement-preserving relations so that they can be defeating relations for other
entitlement-preserving relations. In Chapter V I will adopt the latter approach.
CHAPTER III:

Brandom’s Account of the Appropriate Circumstances for Attributing Inferential Commitments and Why it Does Not Adequately Allow for Defeaters that are not Refuters

PART 1: Defeaters that are not Refuters

I pointed out in Chapter I that defeasible inferential relations must have defeaters which are not also refuters. This is one way to emphasize that an inferential relation’s being defeasible (in the inconclusive sense) is not the same as an inferential relation’s being refutable. This distinction, I briefly argued, needs to be accommodated by any adequate account of defeasible reasoning—and Brandom’s theory is no exception. In this section of this chapter, I will show that even though Brandom is often not careful to distinguish defeasibility from refutability, there are two features of Brandom’s broader account of reasoning that commit Brandom to the claim that scorekeepers can remain implicitly committed to inferential relations that they take to be defeated.

The first feature is that scorekeepers take all of their interlocutors (and themselves) to be rationally bound by the same set of inferential relations. That is, a scorekeeper uses the set of inferential relations that she is implicitly committed to when updating her interlocutors’ scores (rather than the set that her interlocutors are committed
to). Just after presenting his scorekeeping algorithm (discussed at length in the previous chapter) Brandom writes

In this way the broadly inferential content that [S1] associates with [S2]’s claim determines the significance [S2]’s assertional speech act has from the point of view of [S1]’s scorekeeping, the difference it makes to the deontic attitudes of attributing and acknowledging commitments and entitlements by means of which [S1] keeps track of everyone’s deontic statuses.(MIE 191, emphasis added)

The second feature of Brandom’s account that contributes to its needing defeaters that are not refuters is its claiming that even though one interlocutor’s being committed and entitled to a claim $p$ may entitle that interlocutor to another claim $q$, a different interlocutor who is also committed and entitled to $p$ might not be entitled to $q$ because he has an even stronger reason against $q$. For example an interlocutor may be entitled to *Lights* as a result of his being committed and entitled to *Struck*, but another interlocutor who is also committed and entitled to *Struck* might not be entitled to *Lights* because she is also committed to *EMF* which (in context) acts as a stronger reason against *Lights*. This is clearly seen in the defeasibility conditions built into entitlement-preserving relations. “The premises of these inferences entitle one to commitment to their conclusions (in the absence of countervailing evidence) but do not compel such commitment.”(MIE p169).21

---

21 Elsewhere Brandom writes,
If a scorekeeper took an inferential relation to be refuted (and therefore no longer accepted it) in those cases where she takes someone (possibly herself) to be committed and entitled to its premises but not entitled to its conclusion, then that scorekeeper could not use that same inferential relation to update someone else’s score. This would be at odds with the second feature of Brandom’s account which says that an inference may go through for one person but be blocked for another. Therefore, it must be possible for scorekeepers to take an inferential relation to be defeated for someone (and therefore not be obliged to update that person’s score according to it) while maintaining commitment to that inferential relation so that they are obliged to update other peoples’ scores according to it. That is, Brandom’s account requires that some inferential relations can be defeated without being refuted, and therefore that some inferential relations have defeaters which are not also refuters.

As we will see later in the chapter the sorts situations that I am primarily concerned with when it comes to allowing for defeated but unfuted inferential relations are those where a scorekeeper takes a defeating inferential relation to be defeated (but not refuted) for themselves. For example, a scorekeeper should update the scores of her interlocutors according to the inferential relation between EMF and Lights even if that scorekeeper takes that inferential relation to be defeated for herself (perhaps because she is committed to FCage). The remainder of this chapter will argue that Brandom’s account

Endorsing a doxastic inference (one whose premises and conclusions are claims, that is, expressions of possible beliefs), treating that inference as entitlement- or commitment-preserving for one interlocutor, involves treating it as good for all interlocutors—subject, as always, to disqualifications by commitment to incompatible claims, and with the proviso that differences in collateral doxastic commitments can make a difference in what premises are available as auxiliary hypotheses in such inferences. (MIE 250 Italics in original)
of the appropriate circumstances for attributing inferential commitments prevents this, because scorekeepers on Brandom’s account would cease to be implicitly committed to an inferential relation (thereby taking it to be refuted) if they are not disposed to update their own score according to it.

**Part 2:**

**Informal Presentation of Brandom’s Account of the Appropriate Circumstances for Attributing Inferential Commitments.**

Brandom’s account of the appropriate circumstances for attributing inferential commitments is much more complex than it might initially appear. In MIE chapter 3 Brandom presents what I will call the *basic dispositional account* (of attributing inferential commitments). The position endorsed there is, very roughly, that we may appropriately attribute a particular inferential commitment to someone in virtue of how that person is disposed to update score in the reasons game. The complications emerge later in MIE where we see Brandom adopt a sort of Quinean holism where scorekeepers may change their set of inferential commitments whenever they change their assertional commitments. In presenting Brandom’s account I will first explain the basic dispositional account (primarily presented in MIE chapter 3) and then go on to explain the more elaborate holistic details of the account. After informally presenting all these aspects of
Brandom’s position, I will in the next section give a more formal presentation that will show more clearly how Brandom’s basic account and his inferential holism fit together.

**Brandom’s Basic Dispositional Account** (of the appropriate circumstances for attributing inferential commitments).

Because we are supposing (as Brandom does) that the scorekeepers we are interpreting lack the vocabulary to make their inferential commitments explicit, it follows that we cannot attribute inferential commitments to these scorekeepers in virtue of their making (or being disposed to make) assertions expressing inferential commitments. Instead, it is we logically competent theorists who attribute inferential commitments to these scorekeepers in virtue of how they *use* different expressions.

We saw in Chapter I that we can understand Brandom’s scorekeeping vocabulary as the meta-vocabulary he uses for presenting the appropriate circumstances for and consequences of attributing commitments to each of his three types of inferential relations. By understanding content as inferential role and each type of inferential relation as connecting attributions of commitments and entitlements, it becomes possible to understand how social practices of keeping score on commitments and entitlements confer inferential roles articulated along these four dimensions on the expressions they are caught up in them. That is, social deontic scorekeeping provides an explanation of how expressions must be *used* in order to have contents of this sort associated with them—
associated not by the theorist’s stipulation but by the practical attitudes of practitioners whose linguistic conduct is being interpreted. (MIE 190)²²

For example, if a scorekeeper takes an interlocutor’s asserting (and being entitled to assert) \( p \) to entitle that interlocutor to \( q \) then that scorekeeper takes \( q \) to be an entitlement-preserving consequence of \( p \) (MIE 174). Elsewhere in MIE chapter 3 Brandom writes,

Undertaking commitment to \( p \) is undertaking commitment as well to its inferential consequences—to those claims \( q \) that are related to it as conclusions of commitment-preserving inferences having \( p \) as a premise. So if, as a result of \( B \)’s assertion, \( A \)’s deontic attitudes change in that \( A \) comes to attribute to \( B \) a commitment to \( p \), then \( A \) is obliged also to attribute to \( B \) commitment to \( q \). Or rather, \( A \)’s treating this as a good inference consists in \( A \)’s being disposed to keep score in this way, linking the attribution of commitment to \( q \) consequentially to attribution of commitment to \( p \). Again, for \( r \) to be incompatible with \( p \) is for commitment to \( p \) to preclude entitlement to \( r \). For \( A \) to treat these contents as incompatible is for \( A \) to be disposed to withhold attribution of entitlement to \( r \) whenever \( A \) attributes commitment to \( p \). (MIE 186)

These quotations²³ strongly suggest the following account:

---

²² Brandom says four dimensions because he is including the reliability inferences used in perception and action, which I am not addressing.

²³ See also BSD 117-118.
S is implicitly committed to $p \rightarrow q$ (a commitment-preserving relation) IF S is disposed to attribute commitment to $q$ in response to her coming to attribute commitment to $p$.

S is implicitly committed to $p > q$ (an entitlement-preserving relation) IF S is disposed to attribute entitlement to $q$ in response to her coming to attribute commitment and entitlement to $p$.

S is implicitly committed to $p//q$ (an incompatibility relation) IF S is disposed to withhold entitlement to $q$ in response to her coming to attribute commitment to $p$.

It is unlikely that these are the only score updating dispositions that would entitle us theorists to attribute these inferential commitments, but Brandom himself does not give us much more to go on. For example, while Brandom does not say as much, it is probably reasonable to assume that we would also be entitled to attribute to S a commitment to $p \rightarrow q$ if S is disposed to withdraw her attribution of commitment to $p$ in response to her withdrawing her attribution of commitment to $q$ (as if using modus tollens). It is unfortunate that Brandom does not present his account of the circumstances for attributing inferential commitments in nearly as much detail as he does his account of the consequences of attributing (as seen in his scorekeeping algorithm), but thankfully
many of these finer details are not necessary for my argument that Brandom’s account of
defeasible reasoning does not allow for defeaters that are not refuters.

**Brandom’s Inferential Holism**

It is quite tempting to think of a scorekeeper’s inferential commitments, on
Brandom’s account, as being the scorekeeper’s relatively static conception of the “rules”
of the reasons game. This would be in accord with more traditional accounts of meaning
that try to distinguish between the inferential consequences that a claim has essentially in
virtue of its meaning, and the inferential consequences that a claim has only contingently
in virtue of other collateral premises (or what the facts happen to be.) For example, we
may say that we can infer “Jack is male” from “Jack is Tom’s uncle” simply in virtue of
the meanings of these propositions. By contrast, we may say that we can infer “Jack
robbed a store” from “Jack is Tom’s uncle” not because of the meanings of these
propositions, but only because we know some additional premise “Tom’s only uncle
robbed a store”. Brandom, however, rejects this sort of approach.

One privileges *some* of the inferences a concept is involved in as
constitutive of it, treating the rest as warranted by collateral
information…The difficulty faced by this approach is just the one Quine
emphasized: saying what it is about the practices of using expressions that
deserves to be characterized as treating some claims and inferences
involving a concept as essential to it, and others as providing merely
ancillary information about what it applies to. (MIE 484 emphasis in original)

Brandom, following Quine, adopts a sort of holism that blurs the distinction between meanings and beliefs. What inferentially follows from a set of claims largely depends upon what the facts happen to be, and a scorekeeper’s set of inferential commitments largely depends upon what assertional commitments she acknowledges. Furthermore, claims do not have any inferential consequences whatsoever outside of the context of a large set of collateral (or auxiliary) premises. On Brandom’s account, all material inferential relations are contingent on collateral premises. This is clearly seen where Brandom writes:

[W]hat one takes to follow from what (committively or permissively) depends on what collateral premises one is committed or entitled to. This principle applies no less to the inferences incorporated in conceptual contents than those that relate conceptual contents among themselves. To see that the inferential relations internal to a concept are methodologically on par with those external to concepts is the essence of the sort of holism about meaning that Quine introduces in “Two Dogmas of Empiricism,” just as a result of considering the role of auxiliary hypotheses in determining the significance of different commitments…When the inferences implicit in the use of a word are made explicit in the form of

---

24 See Quine (1951) and Quine (1995)
conditionals, the fact that the proprieties governing them are relative to a background of collateral commitments is manifest. Which conditionals one endorses depends in part on what other claims one endorses. When the circumstances and consequences of application potentially associated with a particular word are made propositionally explicit, which conditionals connecting them an interlocutor endorses varies with the other commitments, doxastic and inferential, that the interlocutor undertakes. (MIE p.587)

We see then that a scorekeeper’s set of inferential commitments may change any time she changes her set of assertional commitments. But the holism that Brandom is committed to goes even farther than that. It maintains that we cannot even systematically identify all and only those sets of assertional commitments that a scorekeeper’s being committed to some inferential relation is dependent upon. That is, we cannot even in principle identify rules like ‘Scorekeeper S is committed to inferential relation IRa whenever some particular set of claims A is a subset of S’s assertional commitments. Whether or not a scorekeeper would be committed to IRa were she committed to set of claims A will depend upon whether or not she is committed to various other particular claims that are not members of A. To put it another way, whether or not a particular collateral premise makes a difference to the material goodness of an inferential relation
will depend upon the presence or absence of yet further collateral premises that have thus far not been mentioned.\textsuperscript{25}

This is actually one of the key points that Brandom uses his match-lighting example to illustrate. The appropriateness of one’s inferring \textit{Lights} from \textit{Struck} (defeasibly) depends upon whether or not \textit{EMF} is present as a collateral premise, but whether or not \textit{EMF} makes this difference to the inference will itself depend upon whether \textit{FCage} is present as a collateral premise, and so on. We are never warranted in claiming to have identified all and only those sets of collateral premises in the context of which a particular inference would be good, \textit{such that the presence or absence of further collateral premises would not make a difference}. Recall (from the discussion in Chapter I) how Brandom in the context of his match-lighting example writes

\begin{quote}
The problem is not just that we would need an \textit{infinite} list of the conditions being ruled out—though that is true. It is that the membership of such a list would be \textit{indefinite}: we do not know how to specify in advance what belongs on the list.\textit{(AR 88 italics in original)}.
\end{quote}

The forgoing discussion leads us to a rather radical conclusion: \textit{for any set of claims} \textit{A}, and any claim \textit{x}, there is some set of collateral premises against which \textit{x} would follow from \textit{A} and a different set of collateral premises against which \textit{x} would not follow.

\textsuperscript{25}If one took such dependence relations to be exceptionless then the most that we could say is that a scorekeeper’s having a particular inferential commitment \textit{depends} on the entirety of the particular set of assertional commitments that she takes herself to have—nothing more, nothing less. I, however, am willing to treat such dependence relations as being relative to some background set of collateral premises that we theorists are using, and as therefore refutable (and sometimes defeasible).
from A. If one were to deny this then one would be effectively claiming that there are some inferential relations that do not depend upon what collateral premises are available. That Brandom himself acknowledges the radical conclusion is especially evident in BSD Chapter 3 where he argues that the presence or absence of even absurd gerrymandered properties like ‘fridgeon’ (a particle is a fridgeon just in case Jerry Fodor’s fridge is on) could make a difference to one’s entitlement to any claim. Brandom writes,

For any complex relational property such as being a fridgeon or having old-Provo-coloured eyes, we can describe some inferential circumstances (however outré) in which the credentials of some significant claim would turn precisely on the presence or absence of that property (BSD 80)

Holism doesn’t just tell us that a scorekeeper’s inferential commitments at a given time depend upon some of the collateral premises she is using at that time. Holism tells us that the scorekeeper’s inferential commitments (and the scorekeeping dispositions they depend upon) are wholly indeterminate unless and until we specify (or at least make an assumption about) the entire set of collateral premises that the scorekeeper is using. We may be able to say that a scorekeeper would be committed to some particular inferential relation if she were committed to all and only these assertional commitments. But all bets are off if we are only talking about a proper subset or a superset of the scorekeeper’s assertional commitments—or even if the scorekeeper comes to acknowledge a new assertional commitment. Similarly, we can say things like ‘S would not be committed to inferential relation $IR_a$ were she not committed to $p$', but such statements are only meaningful relative to S’s using an entire set of commitments and entitlements as collateral premises.
One may suspect that Brandom’s holism and its weighty consequences would only apply to commitments to entitlement-preserving inferential relations since they are the only type of inferential relations that Brandom takes to be defeasible.26 This, however, is not the case:

[W]hat a given endorsement of a claim commits one to, is entitled by, and is incompatible with depends upon what else one is committed to, on what collateral information is available as auxiliary hypotheses for the inferences in question.(MIE 477).

While only entitlement-preserving relations are defeasible, the material goodness of any material inferential relation depends upon collateral premises. Inferential holism (at least the sort we are concerned with here) concerns the circumstances in which scorekeepers count as implicitly committed to inferential relations, not the consequences of their inferential commitments. Defeasibility, on the other hand, has to do with the consequences of scorekeepers’ being committed to certain types of inferential relations. Incompatibility relations are indefeasible in that if a scorekeeper were committed to The bird is a swan // The bird is black then she would be obliged to withhold attributing entitlement to ‘The bird is black’ to anyone whom she attributes the commitment ‘The bird is a swan’—regardless of what other commitments and entitlements she attributes to that person. But the scorekeeper’s being committed to that incompatibility relation does not entail the scorekeeper’s being committed to that inferential relation’s being irrefutable—as if it were not contingent on what the facts are. That is, the scorekeeper is

26 Again, I am only focusing on theoretical reasoning and am therefore ignoring the reliability inferences involved in perception and action which are also defeasible.
not committed to never disavowing her commitment to that inferential relation in response to her changing her mind about what the facts are. It is only if we mistakenly treat holism as a thesis about the consequences of one’s being committed to an inferential relation that commitment to commitment-preserving or incompatibility relations appears non-holistic.

For the purposes of this dissertation, the importance of holism is more structural than epistemological. That is, inferential holism is important to our discussion because it describes part of the scorekeeping structure that Brandom claims is necessary for reasoning and using language. It tells us that there is no answer to the question ‘what scorekeeping dispositions does scorekeeper S have?’ except in relation to S’s using all and only the members of some particular set of claims as collateral premises. The present point is not that holism raises some epistemological barrier that makes it difficult for us theorists to interpret some scorekeeper’s updating dispositions from the behavioural evidence available to us (though this might be true).

I will actually be assuming that we theorists already know the various different ways that a scorekeeper would be disposed to update score relative to her using different sets of claims as collateral premises. The central challenge will not consist of trying to discern how a scorekeeper would add and remove commitments and entitlements were she to use various different sets of claims as collateral premises, but figuring out which potential set of collateral premises (or sets of collateral premises) is relevant to our attributing inferential commitments to that scorekeeper.
I have shown that (on Brandom’s account) a scorekeeper’s inferential commitments are largely determined by how that scorekeeper updates score against a large set of background collateral premises, but I have not said much about which sets of commitments and entitlements it is that a scorekeeper uses as collateral premises when assessing what follows from claims. Let us call a set of commitments and entitlements a *score set*. There are infinitely many different score sets that a scorekeeper could use as collateral premises when assessing a claim’s inferential consequences. Even if we knew all the different ways that a scorekeeper would update score were she to use any of the different score sets as collateral premises, we would not know the scorekeeper’s inferential commitments unless we know which particular score set the scorekeeper uses to assess a claim’s inferential consequences. I will call the score set that acts as the collateral premises that a scorekeeper uses to assess a claim’s inferential consequences the *reference score set*. Once we have identified a scorekeeper’s reference score set, we will be able to attribute inferential commitments to that scorekeeper in virtue of how she is disposed to add or remove commitments and entitlements to the reference score set in response to other commitments and entitlements being added to it (or removed from it).

When a scorekeeper S1 is updating interlocutor S2’s score in response to S2’s asserting p, there are at two likely answers to be given to the question, which score set does S1 use as the references score set?: (1) The score set that S1 takes herself to have, (2) The score set that S1 takes S2 to have.
There is no doubt that the assertional commitments and entitlements that S1 attributes to S2 make a difference to how S1 is obliged to update S2’s score. For example, suppose that S1 is implicitly committed to

\[ \{ \text{Fido is warm blooded, Fido is featherless} \} \rightarrow \text{Fido is a mammal} \]

and S2 asserts \textit{Fido is warm blooded}. Whether or not S1 takes S2 interlocutor’s assertion to consequentially commit the interlocutor to \textit{Fido is a mammal} will depend upon whether S1 takes S2 to be committed to \textit{Fido is featherless} (or perhaps various other claims). This may give us the impression that S1’s reference score is whatever score set she takes S2 to have. This however confuses the circumstances for and consequences of attributing inferential commitments. There is no question that whether or not S1’s being obliged to attribute to S2 a commitment or entitlement to an inferential relation’s conclusion largely depends on S1’s attributing to S2 a commitment (and perhaps entitlement) to every premise of that inferential relation (rather than just one of the premises). But this is not the issue that we are presently concerned with. We want to know what inferential commitments we can appropriately interpret S1 as having in the first place—which score set does S1 use as collateral premises when assessing what inferentially follows from the set of claims \{Fido is warm blooded, Fido is featherless\}?

Central to Brandom’s account of objectivity is a distinction that each scorekeeper makes between the set of commitments and entitlements that each interlocutor merely \textit{takes himself or herself to have}, and the set of commitments and entitlements that each interlocutor \textit{actually has} (according to the scorekeeper). For example, a scorekeeper S1 and her interlocutor S2 may both take S2 to be committed to the claim ‘The whale splashed the crowd’. However, S1 might also take S2 to have consequentially undertaken
commitment to ‘A mammal splashed the crowd’ even if S1 does not take S2 to acknowledge this commitment. Brandom writes,

[W]hat someone is committed to may (according to a scorekeeper) not only outrun, but even conflict with, what that interlocutor is prepared to acknowledge. The scorekeeper must keep two sets of books. (MIE 488)

This distinction can exist because (1) each scorekeeper/interlocutor uses her own set of inferential commitments when updating what she takes her interlocutors to be actually committed and entitled to, and (2) each scorekeeper/interlocutor will have a different set of inferential commitments. Thus, one chief reason why S1 and S2 may disagree about what S2 is really committed to as a result of S2’s assertion is that S1 and S2 have different inferential commitments.

One of the key features of Brandom’s account that gives rise to different scorekeepers/interlocutors having different inferential commitments is that they each use a different set of collateral premises when they each assess what actually follows from a claim. What each scorekeeper takes to actually (or objectively) follow from a claim largely depends on what they take to be the facts, and each scorekeeper is going to have a different set of assertional commitments which they will each take to constitute the facts.

The necessity for this dual score follows from the fact that there are in principle two places a scorekeeper can draw auxiliary hypotheses from in extracting the inferential consequences of (and so the commitments consequentially undertaken by) a set of commitments some individual is taken to acknowledge. Those auxiliary hypotheses may be other
commitments the individual acknowledges, or they may be commitments the scorekeeper undertakes (acknowledges), rather than attributing as acknowledged. Since these latter represent the facts (facts being just true claims), according to the scorekeeper, these latter consequences are those that, according to the scorekeeper, actually follow from the claims made (given how things really are), regardless of whether the one making the claims realizes that they follow or not. (MIE488)

We see from this that if we are concerned with what a scorekeeper takes to actually follow from claims (what that scorekeeper’s inferential commitments are) then that scorekeeper’s reference score set consists of the set of commitments and entitlements that she takes herself to have—even when she is updating the score of an interlocutor who she takes to have different commitments and entitlements than she does.

I want to emphasize that this is not an epistemological point, but a structural one. In order for a scorekeeping practice to be linguistic and rational, it must (according to Brandom) have this structure. If we didn’t know what scorekeeping dispositions the scorekeeper has or what score set she takes herself to have, then it would make perfect sense to look at how she would update the scores of her interlocutors in addition to her own, but this is not the scenario we are presently dealing with. I am assuming that we already know how the scorekeeper would be disposed to update score were she to use a wide range of different score sets as collateral premises. Once we know which score set the scorekeeper takes herself to have, then we will know everything that we need to know
about the scorekeeper in order to attribute inferential commitments to her. These inferential commitments will have consequences for how the scorekeeper will be obliged to update the scores of her interlocutors, but we don’t need to consider how the scorekeeper would update her interlocutors’ scores in order to identify these inferential commitments.

However, scorekeepers must also be able to assess what inferential significance a claim has relative to the set of commitments that their interlocutors acknowledge. That is, scorekeepers must have a view on what would inferentially follow from the claims that her interlocutor asserts, were her interlocutor’s view of the facts correct. This is necessary in order for scorekeepers to identify what particular commitment their interlocutor is undertaking when asserting a particular sentence. Because a scorekeeper and her interlocutor have different views on what the facts are, they will use different expressions to identify what exactly the interlocutor is making his assertion about.

Our practical grasp of the objective dimension of conceptual norms...consists in the capacity to coordinate in our scorekeeping the significance a remark has from the perspective of the one to whom the commitment it expresses is attributed and its significance from the perspective of the one attributing it. This requires recognizing the different specifications of the same claim that correspond to extracting its inferential consequences and antecedents in the context of other commitments that are acknowledged as true by the scorekeeper, on the one hand, and extracting them in the context of other commitments acknowledged by the target of that scorekeeping, on the other.(MIE 598)
In light of this we need to be careful to distinguish two importantly different senses in which we can talk about the manner in which a scorekeeper would update some particular score set. Suppose that scorekeeper S uses score set B as collateral premises, takes her interlocutor to have score set A, and takes the same interlocutor to have asserted $p$. When we say “Scorekeeper S would update score set A in manner $M$ (in response to claim $p$ being added to $A$), we might mean either of the following: (1) S would update score in manner $M$ were she to use score set $A$ as collateral premises (rather than score set $B$); (2) Given that S is using score set $B$ as collateral premises, S would update score set $A$ in manner $M$. The first meaning concerns what S takes to be her interlocutor’s perspective on what would follow from $p$. It is also important to our assessing what inferential commitments S would have were she to use score set $B$ as her reference score set. The second meaning concerns how S’s inferential commitments (which depend upon S’s using score set $B$ as collateral premises) would oblige S to update her interlocutor’s score. This is S’s perspective on what her interlocutor’s asserting $p$ really commits and entitles the interlocutor to (regardless of whether the interlocutor herself recognizes it). This has more to do with the consequences of S’s having particular inferential commitments than it does the circumstances for attributing those inferential commitments.

In order to avoid confusion, any time I talk about how a scorekeeper would, or is disposed to, update some score set A (in the context of discussing the circumstances for attributing inferential commitments), I will be treating A as the set of collateral premises that the scorekeeper would be using. When we are assessing a scorekeeper’s inferential commitments, we only need to know how that scorekeeper would add and remove commitments and entitles to her reference score set in response to other
commitments and entitlements being added to or removed from it. We don’t need to know how the scorekeeper would use her reference score set as collateral premises when updating some other score set.27

The finer details of Brandom’s account of how interlocutors are able to secure a common topic of conversation despite their each having a different view of what inferentially follows from claims are very complex and will not be explored here.28 My central focus is the circumstances for and consequences of attributing inferential commitments, which I take to concern the scorekeepers’ perspectives on what claims are actually inferentially related to what others, and what their interlocutors are actually committed and entitled to (even if the interlocutors themselves do not recognize it). I am also assuming that we theorists have already identified and individuated the various contents that scorekeepers attribute commitment and entitlement to.29

---

27 Once again, we are not trying to figure out a scorekeeper’s various updating dispositions relative to her using different sets of collateral premises. I am assuming that we already know this, and are instead trying to figure out how these warrant our attributing different inferential commitments.

28 At least the last third of MIE is largely devoted to solving this problem.

29 Part of the whole strategy that Brandom adopts in MIE is to use the account of how scorekeepers interpret their interlocutors to make sense of what we interpreters are doing when interpreting these scorekeeping practices. Just as scorekeepers individuate contents from their own perspective (which includes what they take the facts to be) when interpreting their interlocutors, so too do we theorists individuate from our own perspective the contents that we interpret scorekeepers to be attributing commitments and entitlements to. We could be wrong about how the scorekeepers we are interpreting individuate contents, but we can’t be too far off otherwise we couldn’t even interpret them as participating in the reasons game.
PART 3: Putting All the Pieces Together:

A More Formal Presentation of Brandom’s Account of the Appropriate Circumstances for Attributing Inferential Commitments

Let us now put all the pieces together. On Brandom’s account a scorekeeper is implicitly committed to a particular inferential relation at a particular time if the scorekeeper at that time is disposed to update her own score in some specified manner (adding and removing commitments and entitlements in response to other commitments and entitlements being added or removed). Furthermore, the manner in which that scorekeeper is disposed to update her own score at a particular time is largely dependent upon the particular set of commitments and entitlements that she takes herself to have at that time (which function as collateral premises).

One of the consequences of Brandom’s holism is that even if we know that scorekeeper S would update score set $A_1, A_2, A_3 \ldots$ in manners $M_1, M_2, M_3 \ldots$, we cannot use this knowledge to generate a set of general rules from which we could then derive the manner in which S would update score $A_n$. Without reference to S’s using some particular score set as collateral premises, there simply is no determinate structure to S’s score updating practices. Therefore, when we attribute inferential commitments to a scorekeeper we are not attempting to posit a set of general rules that we can use to systematically explain why the scorekeeper would update her score in different manners were she to take herself to have different sets of commitments and entitlements.
These considerations may lead one to question what sort of explanatory role inferential commitments can serve. (I suspect that it is just this sort of worry which led Quine to deny that meanings can play a useful explanatory role). To see the explanatory role of inferential commitments in Brandom’s account we need to look at the consequences of attributing them. The point of attributing inferential commitments to a scorekeeper is that they let us systematically explain how that scorekeeper is obliged to update the scores of any of her interlocutors when they make any of the different types of moves in the reasons game (assertions, disavowals, challenges, etc) with different types of claims. Even though a scorekeeper’s having a particular set of inferential commitments at a given conversation stage is relative to the particular set of assertional commitments and entitlements that she takes herself to have at that stage, the consequences of her having these inferential commitments applies very generally to different interlocutors and different claims used in different types of moves.

To highlight and clarify some of the points just made, I will present Brandom’s account of the appropriate circumstances for attributing inferential commitments in terms of a function called \textit{ScoreUpdate}. This function’s arguments consist of a scorekeeper, a score set (acting as a set of collateral premises), and a set of changes that can be made to the score set. The function outputs the score set that results from that scorekeeper’s updating the indicated score set in response to the various changes being made to it. Its syntax is as follows:

\[
\text{ScoreUpdate}(<\text{Scorekeeper}>, <\text{Score Set}>, <\text{Set of Score Changes}>) = <\text{Score Set}>
\]

\footnote{See Quine (1951) and Quine (1995)}
How the \textit{ScoreUpdate} function works and why it outputs the particular score sets that it does is irrelevant. It is intentionally mysterious. Much of the point of appealing to this function is to emphasize that holism makes it impossible for us to appeal to a set of general underlying rules to explain or predict how a scorekeeper would update score were she to use any particular score set as collateral premises.

The score set that is outputted by \textit{ScoreUpdate} when some score set $A$ is inputted into it will typically not be the same as the score set that would result from the scorekeeper’s updating $A$ according to her inferential commitments. Often when a scorekeeper updates score set $A$, her reference score set will be some score set other than $A$. When this happens, the scorekeeper’s inferential commitments (relative to her actual reference score set) will be different from those that she \textit{would} have \textit{were} she to use $A$ as her reference score set. Similarly, a scorekeeper’s inferential commitments relative to some set of collateral premises other than $A$ will oblige the scorekeeper to update score set $A$ in some particular manner in response to certain moves being made in the reasons game, but these inferential commitments do not describe how the scorekeeper would be disposed to update score \textit{were} $A$ her reference score set.

If we wanted \textit{ScoreUpdate} to output the score set that results from the scorekeeper’s using her reference score set as collateral premises when updating some other score set, then the \textit{ScoreUpdate} function would require an additional argument to indicate this other score set. But this would only be useful to our attributing inferential commitments if we were trying to identify a scorekeeper’s score updating dispositions. Once again, the primary concern here is structural rather than epistemological. I am interested in the supposed relationship between score updating dispositions, assertional
commitments, and inferential commitments; and I am assuming that we have already identified the different manners in which a scorekeeper would be disposed to update score were she to use different score sets as collateral premises.

A score set consists of a set of commitments and a set of entitlements, and will be represented using capital letters \( A, B, C \) etc. I will represent a set of commitments containing claims \( p \) and \( q \), and a set of entitlements containing \( q \) and \( r \) as \( C\{p, q\} \) and \( E\{q, r\} \). When I say that some scorekeeper takes someone (possibly herself) to have score set \( A \), I will be assuming that the scorekeeper takes that person to have all and only those commitments and entitlements that are members of \( A \).

Changes to score sets can be represented using the symbols \(+\) and \(-\) to prefix a set of commitments or entitlements. For example, the score set \( A + C\{p\} - E\{q, r\} \) will be the score set that results from adding commitment \( p \) to and removing entitlements \( q \) and \( r \) from score set \( A \). Often, I will use \( +CE\{p\} \) or \( -CE\{p\} \) to indicate the addition or removal of both commitment and entitlement to \( p \). When a score change does not immediately follow some specification of a score set then it is to be treated as a change that could be made to a score set. This is how they will typically appear in \( ScoreUpdate \) functions.

With the notation just described, we can use the \( ScoreUpdate \) function as follows to denote the score set that results from Sally’s updating score set \( A \) in response to commitment and entitlement \( p \) being added and entitlement \( q \) being removed.

\[
ScoreUpdate(Sally, A, +CE\{p\} - E\{q\})
\]
Assuming that Sally takes the indicated score changes to make a difference to whether various other commitments and entitlements should be in the score set, then this score set will not simply consist of \( A + CE\{p\} - E\{q\} \). Sally is likely disposed to add or remove commitments and entitlements to claims other than \( p \) and \( q \). These additional changes are what Sally takes to be the scorekeeping consequences (or scorekeeping significance) of applying the indicated changes to score set \( A \).\(^{31}\) It is in virtue of these that we may attribute inferential commitments to Sally.

In order to more easily denote these scorekeeping consequences, we can use the \textit{ScoreDiff} function. This function takes two score sets as arguments and outputs a set of score changes that must be made to the first score set in order for it to become identical to the second. Its syntax is as follows:

\[
\text{ScoreDiff}(< \text{Score set 1} >, < \text{Score set 2} >) = < \text{Set of score changes} >
\]

For example, if score set \( A = C\{p, q\}E\{p\} \) and score set \( B = C\{p, r\}E\{q, s\} \) then

\[
\text{ScoreDiff}(A, B) = \{+C\{r\}, - C\{q\}, +E\{q, s\}, - E\{p\}\}
\]

By embedding the \textit{ScoreUpdate} function in the \textit{ScoreDiff} function we are able to identify what a scorekeeper takes to be the scorekeeping consequences of some changes being made to a particular score set. The following denotes what a scorekeeper

\(^{31}\) Brandom would call these ‘Inferential Significances’, but I think that this label misleadingly suggests that the changes that a scorekeeper makes to score are the result of her updating according to her inferential commitments. Technically, ‘scorekeeping significance’, on Brandom’s account, covers more than just inferential significance, since it also incorporates interpersonal dimensions to scorekeeping. These finer distinctions are not relevant to the present discussion.
Sally takes to be the scorekeeping consequences of adding commitment and entitlement \( p \) to score set \( A \). (I will be putting the two different score sets being inputted into the \textit{ScoreDiff} function on separate lines in order to make it easier to read).

\[
\text{ScoreDiff}(A + CE\{p\}, \\
\quad \text{ScoreUpdate}(Sally, A, +CE\{p\})
\]

What this gives us are the various changes that Sally would make to score set \( A \) if commitment and entitlement to \( p \) were added to it. That Sally would make these rather than some other changes to score will, of course, depend on various commitments and entitlements being members of the score set acting as collateral premises. But the particular contribution made by \textit{each} of the different members of \( A \) is entirely glossed over and irrelevant to our assessing Sally’s inferential commitments. By focusing solely on the changes made to score, we are able to leave the collateral premises in the background when articulating which particular inferential relations we take scorekeepers to be committed to. So even if Sally’s being committed to a particular inferential relation \( IRa \) depends upon some set of claims \( B \) being a subset of her commitments, few (if any) of the claims in \( B \) will actually be mentioned in \( IRa \)’s premise set or conclusion. (I will say more on this shortly).

Even if we know how a scorekeeper would update score on a lot of different possible score sets (used as collateral premises) in response to a lot of different changes that could be made to those different score sets, we would still not be in a position to identify that scorekeeper’s inferential commitments. We might know the values of each of the following applications of the \textit{ScoreDiff} functions
\[ScoreDiff(A + CE[p]),\]
\[ScoreUpdate(Sally, A, +CE[p])\]

\[ScoreDiff(B + CE[p]),\]
\[ScoreUpdate(Sally, B, +CE[p])\]

\[ScoreDiff(C + CE[p]),\]
\[ScoreUpdate(Sally, C, +CE[p])\]

But, on Brandom’s account not all of these can be relevant to our appropriately attributing inferential commitments to Sally. As discussed towards the end of the previous section, if we want to know what a scorekeeper takes to actually follow from a claim, then we must know what the scorekeeper takes the facts to be. Therefore, if we want to identify what Sally takes to actually follow from \(p\), then we need to single out the scorekeeping significance that Sally takes \(p\) to have relative to the score set that she takes herself to have. So if Sally takes herself to have score set \(A\) and we want to know what she takes to inferentially follow from anyone’s acquiring commitment and entitlement to \(p\), then we would only consider the following scorekeeping significances—the ones that involve the changes that Sally would make to \(A\).

\[ScoreDiff(A + CE[p]),\]
\[ScoreUpdate(Sally, A, +CE[p])\]

Therefore, even if Sally had interlocutors who were committed to score sets \(B\) and \(C\), we would ignore the values of

\[ScoreDiff(B + CE[p]),\]
\[ScoreUpdate(Sally, B, +CE[p])\]

\[ScoreDiff(C + CE[p]),\]
\[ScoreUpdate(Sally, C, +CE[p])\]
The last thing that we need to know in order to attribute inferential commitments are the various score updating dispositions that warrant our attributing the different types of inferential commitments. As mentioned above, Brandom unfortunately does not give very many details as to how a scorekeeper’s attributing (and withdrawing attributions of) commitments and entitlements warrants our interpreting that scorekeeper as being committed to a specific type of inferential relation. The only scorekeeping dispositions he identifies as warranting our attributing commitments to the different types of inferential relations are the ones already presented above.

Again, it is probably safe to assume that there are various other conditions than just these, but at present we don’t have to concern ourselves with them. When these conditions are presented in the way that they are above, they largely ignore (at least gloss over) the complexities introduced by Brandom’s inferential holism. Holism requires that a scorekeeper’s being disposed to add or remove commitments and entitlements in a particular way must be relative to the scorekeeper’s using a particular set of collateral premises. This is exactly what is being captured in the way that the $ScoreUpdate$ and $ScoreDiff$ functions are being used. For example, we can use the following to say that Sally is disposed to add entitlement $q$ in response to adding commitment and entitlement $p$—relative to her using score set $A$ as collateral premises.\(^\text{32}\)

\[
+E\{q\} \in ScoreDiff(A + CE\{p\}, \\
ScoreUpdate(Sally, A, +CE\{p\})
\]

\(^{32}\) Naturally, Sally could not have this scorekeeping disposition if $A$ already included commitment and entitlement to $p$. This is not a problem because Sally’s scorekeeping dispositions are always relative to her using particular a score set as collateral premises.
Thus, if we assume that Sally takes herself to have score set $A$ then we are warranted in attributing to Sally a commitment to the entitlement preserving relation $p > q$. However, even if we do not know what particular score set Sally takes herself to have; we can still usefully specify the conditions in which it would be appropriate to attribute inferential commitments to her as follows.

Sally is committed to $p > q$ IF

$$A = \text{Whatever particular score set Sally takes herself to have}$$

$$+E\{q\} \in ScoreDiff(A + CE[p],$$

$$ScoreUpdate(Sally, A, +CE[p])$$

I will now present Brandom’s account of the conditions for attributing inferential commitments much more generally for any scorekeeper $S$.

**Commitments to commitment-preserving inferential relations:**

1) For any set of claims $X$ and any claim $y$

2) Scorekeeper $S$ is implicitly committed to $X \rightarrow y$ IF

3) $A = \text{Whatever particular score set } S \text{ takes herself to have}$

4) $+C\{y\} \in ScoreDiff(A + C\{X\},$

$$ScoreUpdate(S, A, +C\{X\})$$

Line 4 clearly does most of the work here. It says that if scorekeeper $S$ were to update score set $A$ in response to commitments to the members of $X$ being added to it then she would add commitment $y$ to that score set (among other changes that she might make). This identifies the scorekeeping significance that $S$ takes $+C\{X\}$ to have against the set of collateral premises $A$. Line 3 simply says that the particular score set being updated in line 5 (whatever its members might be) is the one the scorekeeper takes herself to have (and is therefore the reference score set for $S$). As can be seen below, the
appropriate conditions for attributing commitment to incompatibility and entitlement-preserving relations follow the same pattern except that they involve different changes being made to the reference score set.

**Commitments to Incompatibility Relations**

1) For any set of claims $X$ and any claim $y$
2) Scorekeeper $S$ is implicitly committed to $X // y$ IF
3) $A = \text{Whatever particular score set } S \text{ takes herself to have}$
4) $-E\{y\} \in \text{ScoreDiff}(A + C\{X\},$
   \hspace{1cm} $\text{ScoreUpdate}(S, A, +C\{X\})$)

**Commitments to Entitlement-Preserving Relations**

1) For any set of claims $X$ and any claim $y$
2) Scorekeeper $S$ is implicitly committed to $X > y$ IF
3) $A = \text{Whatever particular score set } S \text{ takes herself to have}$
4) $+E\{y\} \in \text{ScoreDiff}(A + CE\{X\},$
   \hspace{1cm} $\text{ScoreUpdate}(S, A, +CE\{X\})$)

For the argument being developed here, we need not be especially concerned about which specific changes to score warrant attributing commitment to which type of inferential relation. I strongly doubt that Brandom would take these to be the *only* score updating dispositions that would entitle us to attribute each type of inferential commitment. What I hope these different conditions show is the relationship, on Brandom’s account, between the inferential commitments that we may attribute to a scorekeeper and the assertional commitments and entitlements that the scorekeeper takes herself to have. In particular, they show that a scorekeeper’s being committed to a particular inferential relation depends upon the score set that the scorekeeper takes herself to have, and that the particular contribution made by each of the different claims in the
score set is irrelevant. They also show us that Brandom accommodates holism by relativizing a scorekeeper’s inferential commitments to the particular score set that she takes herself to have at a given time, not by trying to incorporate the scorekeeper’s score set into the inferential relations that the scorekeeper is committed to. We say that scorekeeper S is committed to \( p \rightarrow q \) (while she takes herself to have score set \( A \)). We don’t say scorekeeper S is committed to \( \{ A \cup p \rightarrow q \} \). The difference between these is significant when we remember that scorekeepers on Brandom’s account update their interlocutors’ scores according to their own inferential commitments. While many of a scorekeeper’s interlocutors may have \( p \) in their set of assertional commitments, almost none will be committed to all the members of \( A \) (which is every claim that the scorekeeper takes herself to be committed to).

**PART 4: Why Brandom’s Account of the Circumstances for Attributing Inferential Commitments does not allow for Defeaters that are not Refuters.**

At the beginning of this chapter I argued that Brandom’s account of reasoning requires that defeasible inferential relations can have defeaters which are not refuters (and that therefore Brandom’s account requires inferential relations that are defeasible in the inconclusive sense). If Brandom’s account does not allow this, then scorekeepers will not be obliged to update their interlocutors’ scores according to any inferential relations that the scorekeepers take to be defeated for themselves. I will now argue that one of the consequences of Brandom’s account of the appropriate circumstances for attributing
inferential commitments is that it will not in fact allow for defeaters that are not also refuters.

Presumably, one of the ways in which a scorekeeper S can count as implicitly taking some claim \( p \) to be a refuter for some inferential relation \( IRa \) is if S’s commitment to \( IRa \) is contingent on S’s not taking herself to be committed to \( p \) (or if S’s not being committed to \( IRa \) is contingent on S’s being committed to \( p \)).\(^{33}\) This gives us our first premise:

Premise 1:

\[
\text{IF(IF S were to take herself to be committed to } p \text{ THEN the conditions for attributing inferential commitment } IRa \text{ to S would not be satisfied) THEN S takes } p \text{ to be a refuter for } IRa
\]

The second premise concerns the consequences of a scorekeeper taking a claim to be a defeater for an inferential relation. Suppose we assume that in whatever example of defeasible reasoning we are working with, scorekeeper S should take claim \( p \) to be a defeater for inferential relation \( IRa \). Let us also assume that S correctly updates her score with respect to this. It follows that were S to take herself to be committed to \( p \), then S would not take the addition or removal of \( IRa \)’s premises to or from her score set to have any impact on \( IRa \)’s conclusion’s being added to or removed from her score set.

Premise 2:

\[
\text{IF S takes } p \text{ to be a defeater for } IRa \text{ THEN (IF S were to take herself to be committed to } p \text{ THEN S would not take the addition or removal of } IRa \text{'s premises}
\]

\(^{33}\) These contingency relations will be relative to all the other commitments and entitlements that S takes herself to have at that time and they are therefore refutable (and perhaps defeasible).
to or from her score set to have any scorekeeping significance with regards to

*IRa’s* conclusion.)

I take premises 1 and 2 to be relatively uncontroversial (at least to those sympathetic to the sort of project Brandom is undertaking). The third premise, however, is more specific to Brandom’s account of the conditions for attributing inferential commitments. As explained above, we may attribute an inferential commitment to a scorekeeper in virtue of the way that the scorekeeper is disposed to add or remove commitments and entitlements to or from the score set that she takes herself to have (in response to other claims being added or removed). If in response to claim *p*’s being added to or removed from her commitments or entitlements the scorekeeper is not disposed to add or remove claim *q*, then that scorekeeper is not committed to *p* and *q* being inferentially related to one another. For the argument at hand it does not matter which particular patterns of score changes warrant our attributing commitments to which particular types of inferential relations.

Premise 3:

IF S would not take the addition or removal of *IRa’s* premises to or from her score set to have any scorekeeping significance with regards to *IRa’s* conclusion THEN the conditions for attributing inferential commitment *IRa* to S would not be satisfied.

Most of this chapter (particularly parts 2 and 3) has been aimed at providing an explanation and argument for Brandom’s being committed to premise 3. The manner in
which the scorekeeper is disposed to update the score set she takes herself to have is not simply one of various pieces of the evidence available to us when we attribute inferential commitments to that scorekeeper. On Brandom’s account, the manner in which a scorekeeper is disposed to update her own score (in response to different changes made to it) constitutes that scorekeeper’s one and only perspective on what actually follows from different claims. If the scorekeeper is not disposed to update the score set that she actually takes herself to have according to IRa, then she is simply not committed to IRa.

Putting the premises altogether, we get:

Premise 1:

IF(IF S were to take herself to be committed to \( p \) THEN the conditions for attributing inferential commitment \( IRa \) to S would not be satisfied) THEN S takes \( p \) to be a refuter for \( IRa \)

Premise 2:

IF S takes \( p \) to be a defeater for \( IRa \) THEN (IF S were to take herself to be committed to \( p \) THEN S would not take the addition or removal of \( IRa \)’s premises to or from her score set to have any scorekeeping significance with regards to \( IRa \)’s conclusion.)

Premise 3:
IF S would not take the addition or removal of $IRa$’s premises to or from her score set to have any scorekeeping significance with regards to $IRa$’s conclusion THEN the conditions for attributing inferential commitment $IRa$ to S would not be satisfied.

Premise 4 (From 2 and 3):

IF S takes $p$ to be a defeater for $IRa$ THEN (IF S were to take herself to be committed to $p$ THEN the conditions for attributing inferential commitment $IRa$ to S would not be satisfied)

Intermediate Conclusion (From 1 and 4):

IF S takes $p$ to be a defeater for $IRa$ THEN S takes $p$ to be a refuter for $IRa$.

What this argument shows is that if premises 1 and 2 are correct then a consequence of premise 3 is that scorekeepers cannot take a claim to be a defeater without also taking it to be a refuter. Because premise 3 is a consequence of Brandom’s account of the appropriate circumstances for attributing inferential commitments, it follows that scorekeepers on Brandom’s account cannot take a claim to be a defeater without also taking it to be a refuter. I will now apply this argument to the match-lighting scenario.

Let us call the inferential relation between $EMF$ and $Lights IRa$. For the moment we need not concern ourselves with what type of inferential relation $IRa$ is, though we will assume that it is defeasible. In the match-lighting example, the claim $FCage$ is a
defeater for \textit{IRa} and we will assume that our scorekeeper S will treat it as such. What we want to know is whether S’s taking \textit{FCage} to be a \textit{defeater} for \textit{IRa} would, as a result of Brandom’s account, have S take \textit{FCage} to be a \textit{refuter} for \textit{IRa}.

On Brandom’s account S is only going to take \textit{FCage} to be a defeater or a refuter for \textit{IRa} relative to the particular score set that she takes herself to have at that time. Let us assume that S takes herself to have score set \( A \), where

\[
A = C\{Struck, \ldots\}E\{Struck, Lights, \ldots\}
\]

I use ‘\( \ldots \)’ to acknowledge that S is committed and entitled to many more claims than those that are relevant to the example. Assuming that S is disposed to update score correctly, then the following would hold.

\[
-E\{Lights\} \in \text{ScoreDiff}(A + C\{EMF\}, \text{ScoreUpdate}(S, A, +CE\{EMF\}))
\]

It is in virtue of S’s being disposed to remove entitlement \textit{Lights} from her score set in response to commitment and entitlement to \textit{EMF} being added, that we can attribute to S a commitment to \textit{IRa}. Again, we need not concern ourselves with the particular details about what type of inferential relation \textit{IRa} is.

Now let us consider how S \textit{would} update her score \textit{were} S to take herself to have a score set \( B \), where \( B \) is identical to \( A \) except that it also includes commitment and entitlement to \textit{FCage}.

\[
B = A + CE\{FCage\} = \{C\{Struck, FCage, \ldots\}E\{Struck, FCage, Lights, \ldots\}\}
\]

\[
-E\{Lights\} \notin \text{ScoreDiff}(B + C\{EMF\},)
\]

The point that I want to drive home here is that if S takes $FCage$ to be a defeater for $IRa$ then S would not take $+CE\{EMF\}$ to have any scorekeeping significance with regards to $Lights$. Were S to take herself to be committed to $FCage$. Because S’s being committed to $IRa$ depends upon her taking its premises to have some scorekeeping impact on its conclusion (relative to the particular score set she takes herself to have), it follows that if S takes $FCage$ to be a defeater for $IRa$, then S will also take $FCage$ to be a refuter for $IRa$. Or to put it slightly differently, on Brandom’s account there is no difference between a scorekeeper taking an inferential relation to be defeated for herself and her taking it to be refuted.

The negative consequences of S taking $FCage$ to be a refuter for $IRa$ becomes apparent when S actually does take herself, but not her interlocutor, to be committed to $FCage$. Ideally, if the scorekeeper took her interlocutor to be committed and entitled to $Struck$ and $EMF$, then the scorekeeper would not take her interlocutor to be consequentially entitled to $Lights$. When the scorekeeper updates her interlocutor’s score, she should treat $IRa$ as an undefeated defeating relation for the inferential relation from

\[
ScoreUpdate(S, B, +CE\{EMF\}))
\]

\[
+E\{Lights\} \not\in ScoreDiff(B + C\{EMF\}, ScoreUpdate(S, B, +CE\{EMF\}))
\]

\[
-C\{Lights\} \not\in ScoreDiff(B + C\{EMF\}, ScoreUpdate(S, B, +CE\{EMF\}))
\]

\[
+C\{Lights\} \not\in ScoreDiff(B + C\{EMF\}, ScoreUpdate(S, B, +CE\{EMF\}))
\]

---

34 Recall from chapter 1, that $IRb$ is a defeating inferential relation for another inferential relation $IRc$ in virtue of $IRb$’s being responsible for some claim’s being a defeater for
Struck to Lights, even though IRA is defeated for the scorekeeper herself. But if the scorekeeper takes IRA to be refuted by her own commitment to FCage then she will not be obliged to update her interlocutor’s score according to it and will therefore take her interlocutor to be entitled to Lights even though the interlocutor is committed and entitled to EMF, but not FCage.

What all this shows is that Brandom’s circumstances for and consequences of attributing his different types of inferential commitments incorporate materially bad inferences. We started off assuming that a scorekeeper with a particular score set would update her own score correctly. We then used Brandom’s account to identify what inferential relations this scorekeeper would be committed to (or, more importantly, what inferential relations she would not be committed to). We then looked at the consequences of her having (or not having) those inferential commitments and saw that she would in certain circumstances be obliged to update her interlocutors’ scores incorrectly.

**PART 5: Two Unsuccessful Quick Fixes:**

Before moving onto the next chapter I will address two responses that are likely to be raised against my argument that Brandom’s account cannot allow for defeaters that are not also refuters. The responses that I will be considering here are not so much attempts to show that my argument is unsound, but attempts to show that the consequences of the argument can be easily avoided with a few quick fixes or slight reinterpretations of Brandom’s position.

IRAc. IRb would not itself be a defeater for IRAc. In the example at hand, IRAa is a defeating relation (but not itself a defeater) for the inference from Struck to Lights.
The key premise in the argument given above is (roughly) that scorekeepers on Brandom’s account would not count as being committed to an inferential relation $IRa$ if, relative to the particular score set that they take themselves to have at that time, they are not disposed to update score according to $IRa$. The problem is that if the score set that the scorekeeper takes herself to have contains a commitment to a defeater for $IRa$, then that scorekeeper will not update her score according to $IRa$. The quick fixes that I will now consider are attempts to modify the key premise by reconsidering what we should use as the reference score set when attributing inferential commitments.

**Quick Fix 1: Score Sets that a Scorekeeper *Could* take Herself to Have.**

The first quick fix acknowledges that the scorekeeper will not be disposed to take $EMF$ to have any scorekeeping impact on $Lights$ while she is committed to $FCage$. It however, maintains that we can still interpret that scorekeeper as being committed to an inferential relation between $EMF$ and $Lights$ because the scorekeeper *would* take $EMF$ to have a scorekeeping impact on $Lights$ *were* she not committed to $FCage$. So in addition to the inferential commitments that we can attribute to the scorekeeper in virtue of how she *is* disposed to update the score set that she *actually* takes herself to have, we can also attribute inferential commitments to the scorekeeper in virtue of how she *would* update a score set that she *could* take herself to have.

**RESPONSE:**

It is no doubt true on Brandom’s account that there is some score set other than the one that $S$ *actually* takes herself to have such that $S$ *would* update her score according
to \textit{IRa} were \( S \) to take herself to have this other score set. But given inferential holism, this is rather trivial. If we were entitled to attribute to \( S \) a commitment to a particular inferential relation because \( S \) would update her score according to it \textit{were} she to take herself to have a different score set \textcolor{red}{THEN} we could attribute to \( S \) commitment to every possible inferential relation. For any possible inferential relation \( IRx \), there is some score set \( Y \) such that \( \text{were} \) \( S \) to take herself to have \( Y \) \textcolor{red}{THEN} \( S \) would be disposed to update her score according to \( IRx \) (thereby entitling us to attribute commitment to \( IRx \)). There is nothing in principle wrong with assessing a scorekeeper’s inferential commitments by considering how she would update a score set that is not the one that she actually takes herself to have. However, if we do not want attributions of inferential commitments to become vacuous then we must say something about why we are using one score set rather than any other possible score set as the reference score set.

\textbf{Quick Fix 2: “Normal” Reference Score Sets}

Instead of considering how a scorekeeper would update just any score set that she \textit{could} take herself to have, the next quick fix suggests that we only consider \textit{normal} score sets. The idea is that defeaters (for defeasible inferential relations) are deviations from normal conditions and when we attribute inferential commitments to a scorekeeper we should only consider how she would update a normal score set (or what she would take to be a normal score set). Therefore, the reference score set that we should use to assess a scorekeeper’s inferential commitments should not (necessarily) consist of the score set that the scorekeeper actually takes herself to have, but the score set that she takes herself to have minus any commitments (and perhaps entitlements) to defeaters. This would
allow us to interpret a scorekeeper as being committed to an inferential relation and one of its defeaters, because even though the scorekeeper would not be disposed to update her own score set according to the inferential relation, her being committed to that inferential relation would not depend on this.

RESPONSE:

The problem with Quick Fix 2 is that prior to identifying a reference score set, it is a hopeless task to try and distinguish those claims that a scorekeeper takes to be a defeater from those that she does not. First of all, whether or not a scorekeeper takes a particular claim to be a defeater for a particular inferential relation will be (largely) contingent on the particular set of collateral premises that she is using. As a result, we would need to identify a reference score set in order to determine whether a particular claim should be a member of the reference score set. We could stipulate which claims are defeaters and which are not without considering the scorekeeper’s attitudes and updating dispositions, but doing so would bring us too dangerously close to doing the sort of stipulative semantics that Brandom wants to avoid.

To make matters more complicated, we would also need some story about how we are able to distinguish a scorekeeper’s taking some claim to be a defeater for an inferential relation and her taking that claim to be an abnormal refuter for that inferential relation. Presumably, some refuters would not be among whatever set of claims are considered “normal”, but these should not be removed from the reference score set in the way that defeaters are. Otherwise, a scorekeeper’s taking herself be committed to such refuters would not have any impact on her inferential commitments.
The problem with the sort of quick fixes being considered is that they mistakenly fall back on the assumption that there is a determinate rational structure that underlies scorekeeping practices and is independent of any particular set of collateral premises being used in the background. The motivation behind quick fix 1 seems to be the assumption that a scorekeeper’s set of inferential commitments are general rules that purport to express this determinate structure. We would then appeal to these inferential commitments to explain how or why a scorekeeper would update score in a particular manner if she were to use a particular set of collateral premises. It is on this assumption that we are supposed to be able to attribute inferential commitments (as general rules) to a scorekeeper in virtue of how she would update various score sets other than the particular one that she actually takes herself to have. Quick fix two assumes that, independently of the scorekeeper using any particular set of collateral premises, there is a determinate distinction between the set of claims that a scorekeeper takes to defeat a particular inferential relation and the set of claims that she does not take to defeat it.

The problem that holism raises for the two quick fixes is precisely that a scorekeeper’s updating practices have no determinate structure except in relation to the scorekeeper using a particular score set as a set of collateral premises.\(^{35}\) We therefore cannot specify the score set that we should use to assess a scorekeeper’s inferential commitments by appealing to how that scorekeeper would update score generally—indeed, independently of her using any particular score set as collateral premises.

\(^{35}\) It is for precisely this reason that Davidson says that we can’t hope to be able to interpret what someone’s expressions mean without assuming at the outset that that person has most of the same beliefs that we do.
Despite this, the general strategy behind the two quick fixes is a good one: If we change the reference score set to be different from the score set that the scorekeeper takes herself to have then the scorekeeper could be committed to an inferential relation while taking it to be defeated for herself. I think that the most promising way to overcome some of the challenges raised by holism is for us to introduce some additional features into scorekeeping practices which will help us limit the range of possible sets of collateral premises that we have to consider when we attribute inferential commitments. This is the approach that will be adopted in Chapter 6 where I propose a revised account of the appropriate circumstances for attributing inferential commitments—one that will allow for defeaters that are not also refuters.
CHAPTER IV:
Some Pessimism about the Prospects of Incorporating Features From Prominent Accounts of Defeasible Reasoning.

PART 1: The Challenges of Jury-rigging

Review of the Problems

In the previous two chapters I argued that Brandom’s account of defeasible reasoning encounters two different problems, and that therefore his account of the circumstances for and consequences of attributing commitments to the different types of inferential relations must be revised. In Chapter II I argued that Brandom’s account of the consequences of attributing inferential commitments does not allow entitlement-recovery. I showed that no matter what combination of inferential commitments we attribute to a scorekeeper, she will be obliged to update score incorrectly in any instance of defeasible reasoning where someone should be able to gain, lose, and reacquires entitlement to some claim in response to his successively acquiring commitment (and entitlement) to various other claims.

In Chapter III I argued that Brandom’s account of the circumstances for attributing inferential commitments does not allow for defeaters that are not refuters. I showed that if a scorekeeper took herself to be committed to a defeater for an inferential relation and was disposed to update her score correctly then she would no longer be committed to that inferential relation, thereby taking the claim to be a refuter for that
inferential relation. The consequence of a scorekeeper’s taking every claim that is a
defeater for an inferential relation to also be a refuter for the inferential relation is that the
scorekeeper would not be obliged to update her interlocutors’ scores according to any
inferential relation that the scorekeeper takes to be defeated for herself (even if her
interlocutors are not committed to any defeaters for the inferential relation).

**Why Brandom’s Theory is Worth Salvaging**

At this point in the discussion one might treat the arguments given in the previous
two chapters as reasons for giving up on Brandom’s entire account of language and
reasoning. This may be especially tempting for anyone who agrees that most of our
reasoning is defeasible. If we were to assess the merits of Brandom’s overall theory on
the basis of how well it handles defeasible reasoning then I think that we should indeed
reject Brandom’s theory. Similarly, someone who is not already sympathetic to
Brandom’s project would probably treat the arguments that I have given as further
reasons for rejecting Brandom’s position.

My own view is that Brandom’s overall theory is worth trying to salvage. At the
very least, I think that Brandom’s account makes a large step in good direction: it
remains the best worked out account of how the meanings of expressions are determined
by their use; it does not treat representation as a primitive (and therefore does not leave it
mysterious); it is less intellectualist than most other theories (though still more
intellectualist than I would like); it does not depend on naturalism; and it remains
committed to the crucial importance of clarity in philosophical inquiry—the chief virtue
of the analytic tradition. No doubt there will be disagreement about whether many of
these features are strengths or weaknesses. It is well beyond the scope of this dissertation to settle these matters. The point is that if one does value these features (and I think that there are many philosophers that do) then one is likely to see Brandom’s theory as taking a large step in a good direction and worth salvaging from the arguments that I have raised against it.

Even if one thinks that Brandom’s account is a big step in the wrong direction, one should still be able to appreciate it as a detailed and plausible alternative to the sort of representationalism that dominates analytic philosophy. We often gain a lot of philosophical insight into an issue simply by realizing that the way we usually think about things is not the only way to think about them. When we philosophically engage with views that differ from our own we are often led to carefully articulate and critically scrutinize our own assumptions in ways that we might not have otherwise done. Assuming that there is something philosophically valuable about critically scrutinizing our assumptions (or at least certain assumptions), it follows that there is something valuable in engaging with views that differ from our own.\textsuperscript{36} So even if one is not inclined to accept Brandom’s theory, one can at least accept that it makes an important contribution to the philosophical community and that it is therefore worthwhile (up to a certain point) to try to solve some of the problems it faces.

\textsuperscript{36} I am not endorsing the absurd view that every idea makes an important contribution to the philosophical community. Some theories and ideas deserve to be explored more than other ones do. I will not try to identify every feature of theories that make them worth exploring. My point is more that we, as philosophers, should encourage the development of theories that differ from our own (and in particular differ from those that are more dominant). For even if those theories are wrong they can help provide us with useful a new perspective on whatever theory we think is more promising.
What to Expect if We Try to Draw From Other Theories of Defeasible Reasoning

Brandom is by no means the first theorist to discuss defeasible reasoning. There is a lot of literature about defeasible reasoning in the fields of epistemology\textsuperscript{37} and in artificial intelligence.\textsuperscript{38} One might therefore suppose that some of the theories in epistemology and AI have successfully developed accounts of defeasible reasoning that overcome the problems Brandom’s account faces, and that we could perhaps incorporate these accounts (or portions of them) into Brandom’s own theory of language and reasoning.

While I will not argue that there are no good theories of language and reasoning developed in AI or epistemology that can be imported into Brandom’s account of reasoning, I will argue that it is not a very promising direction to take in looking for a solution to the problems raised. I will not argue this on the grounds that epistemologists or AI theorists are unlikely to have developed plausible accounts of defeasible reasoning. What I will argue is that most of the research on defeasible reasoning presumes a philosophical framework that is at odds with Brandom’s broader theory of language and reasoning, and therefore cannot be easily imported into Brandom’s account without seriously changing the nature of his project. To analogize theories of reasoning with machines, Brandom is building his theory of reasoning in such a way that makes it very difficult to incorporate the mechanisms used in most people’s accounts of defeasible reasoning because these mechanisms were designed to fit a different kind of machine. It would be great if we could incorporate components from other people’s theories into

\textsuperscript{37} See Pollock 1987.
\textsuperscript{38} See Reiter 1987 and Alchourron, Gärdenfors, and Makinson 1985.
Brandom’s theory so that it could adequately account for defeasible reasoning, but making it so that these components would work in Brandom’s theory of reasoning is itself a very difficult task. We would be no closer to, if not farther from, an acceptable solution than we were before.

Most of the tradition in which defeasible reasoning is discussed, and in fact most of the analytic philosophy generally, is what Brandom calls representationalist. The central concept in such accounts of reasoning is representation. The meaning of expressions in these sorts of accounts is typically understood as being what the expressions represent (or purport to represent). The goal of reasoning is largely understood to be successful representation and it is in terms of correct representation that truth and correct inference are characterized. Beliefs are usually taken to be mental representations, and something’s having a mind is often equated with something’s representing the world to itself.

Part of what makes Brandom’s general theory of reasoning so interesting is that it is so well worked out but also so very different from the more traditional representationalist theories. But it is these sorts of differences that make it very difficult to import the mechanisms used in other accounts of defeasible reasoning into Brandom’s broader theory of reasoning. I cannot of course give a precise account of what all those difficulties would be because they would vary with the different accounts of defeasible reasoning that we would try to use. What I will do is provide some examples of the challenges we would encounter if we were to try and repair Brandom’s account of reasoning by importing either of two popular accounts of defeasible reasoning.
As I present some of the accounts of defeasible reasoning put forward by authors other than Brandom, I will attempt to use the expression ‘defeat’ (and its associated expressions) in the way that those authors do, rather than in my preferred way (where I distinguish ‘defeat’ from ‘refute’).

**PART 2: John Pollock’s “Defeasible Reasoning”**

**Presentation of the position:**

The first position I will consider incorporating is found in John Pollock’s paper “Defeasible Reasoning”. Pollock’s concern in this article is primarily epistemological. His aim is to computationally characterize “the rules of reasoning”.

Reasoning is guided by rules, and when the reasoning is in accordance with the rules the resultant beliefs are said to be *justified*… The possession of such internalized rules constitutes procedural knowledge. Jointly, these rules comprise a production system. These rules are the rules for “correct reasoning” – the very rules that are the subject of epistemology. (Pollock 483 italics in original)

For Pollock, reasoners have a foundational set of states (or an *epistemic basis*) from which they reason and form new beliefs. These foundational states primarily consist of perceptual states (such as being appeared to as if p), and memory states (such as recalling p) from which we can have defeasible and conclusive reasons for gaining and
losing beliefs. When reasoning leads to the formation of a new belief the reasoner keeps track of the argument used; this consists of the inference rule used in the reasoning and the earlier state (eg belief, perceptual, or memory state) from which the inference is made. What a reasoner can be justified in believing is relative to his or her epistemic basis which grounds the knowledge. (Pollock p.490)

Many of the reasons (arguments) someone can have for a belief are defeasible. A prima facie (or defeasible) reason is one that has potential defeaters. Pollock characterizes defeaters as follows:

R is a defeater for P as a prima facie reason for Q if and only if P is a reason for S to believe that Q and R is logically consistent with P but (P&R) is not a reason for S to believe Q. (Pollock 484 Italics in original)

Pollock then goes on to point out that there are two different kinds of defeasors:

R is a rebutting defeater for P as a prima facie reason for Q if and only if R is a defeater and R is a reason for believing \(~Q\).

R is an undercutting defeater for P as a prima facie reason for S to believe Q if and only if R is a defeater and R is a reason for denying that P wouldn’t be true unless Q were true. (Pollock 485 Italics in original)

Rebutting defeasors are reasons for denying the conclusion of the defeated argument. For example, suppose we took our belief “The bedroom light switch is flipped to the ‘on’ position” to be a good prima facie reason for believing “The bedroom light is
on”. When we then acquire the belief “The power to the house is disconnected” we find ourselves with good reason for denying the conclusion “The bedroom light is on” and thus defeat our original prima facie reason.

Undercutting defeasors need not involve a denial of the conclusion of a prima facie argument, but can instead deny the connection between the premises and conclusion of the argument. For example the claim “The Bedroom light switch is broken” does not act as a reason for denying the conclusion of the argument above—“The bedroom light is on”; instead it amounts to a denial of the connection between “The bedroom light switch is flipped to the ‘on’ position” and “The bedroom light is on”. Undercutting defeaters allow one to deny that an argument (reason) is good without going so far as denying the conclusion of the argument.

A (belief in a) proposition is warranted only if there is some ultimately undefeated argument leading from the foundational states to that proposition. What Pollock means by ‘ultimately undefeated arguments’ can be seen when he writes:

Merely having an argument for a proposition does not guarantee that the proposition is warranted, because one might also have arguments for defeaters for some of the steps in the first argument. Iterating the process, one argument might be defeated by a second, but then the second argument could be defeated by a third thus reinstating the first and so on. A proposition is warranted only if it ultimately emerges from this process undefeated.(Pollock 491)³⁹

³⁹ Pollock provides a much more technical characterization of what ‘ultimately undefeated’ means, but such details need not concern us here.
Very roughly an argument from $p$ to $q$ is ultimately undefeated so long as there are no undefeated arguments supporting either $\sim(p \rightarrow q)$ or $\sim q$. If the argument from $p$ to $q$ is ultimately undefeated then $q$ is warranted. Beliefs are only warranted if an uninterrupted chain of reasoning (or argumentation) can be traced from those beliefs back to the foundational states. The chain of reasoning is interrupted if one or some of the steps in the reasoning (for example from $p$ to $q$) is defeated by a different chain of argument leading from the foundational states to a proposition that rebuts the conclusion of the argument (eg. $\sim q$) or undercuts the relation between premises and conclusion (eg $\sim(p \rightarrow q)$). At one time proposition $q$ might be warranted on the basis of an ultimately undefeated argument leading from $p$ to $q$, but if one later acquires ultimately undefeated arguments in support of $\sim p$ or $(\sim p \rightarrow q)$ then $q$ ceases to be warranted.

**Difficulties Incorporating features of the Account:**

Because the conclusions that can be argued from one set of claims cannot necessarily be argued from a superset of it, the reasoning in Pollock’s account is nonmonotonic (in the technical sense). And, because undercutting and rebutting defeasors can themselves be defeated, Pollock’s account of reasoning allows entitlement recovery.

Furthermore, reasoners on Pollock’s account can very easily take some inferential relations (or arguments) to be defeated for themselves without also taking them to be refuted. Even if a reasoner takes the inference from $p$ to $q$ to be defeated by a newly acquired belief, they can nonetheless remain committed to the inferential relation in virtue of having a conditional expressing it remain in their set of beliefs.
Even though Pollock’s account of reasoning allows entitlement recovery, and allows for defeaters that are not refuters, it has various features that make it very difficult to integrate it into Brandom’s general theory of language and reasoning. For starters, Pollock’s account of defeasible reasoning assumes a sort of foundationalism that Brandom rejects. On Pollock’s account it is necessary that reasoning has a justificatory bedrock so that reasoners can determine which arguments are ultimately undefeated and which are not.\(^4\) Pollock also presumes that if one does not have an argument for a belief \(q\) then that belief is unwarranted. For example when explaining how a reasoner is to update beliefs after having an argument from \(p\) to \(q\) undercut Pollock says that the reasoner would come to believe \(\neg(p \rightarrow q)\) and then remove belief \(q\) if there are no other arguments supporting \(q\) (Pollock p.512). This is also seen in Pollock’s treating one’s having justification for (or being entitled to) \(p\) as one’s having an argument for \(p\).

On Brandom’s default and challenge account of entitlement someone can be entitled to a commitment (belief) without being able to provide reasons for it. But someone’s having default entitlement to a commitment does not mean that one’s entitlement to that commitment cannot be challenged; it just means that any challenge made against one’s entitlement to that commitment must itself be justified.

Claims such as “There have been black dogs” and “I have ten fingers” are ones to which interlocutors are treated as prima facie entitled. They are not

---

\(^4\) More accurately this is required of an ideal reasoner with unlimited time and memory. Pollock holds that humans do not have the computational resources necessary to meet this ideal and he therefore proposes a set of belief updating rules that approximates ideal reasoning. While these updating rules do not require that reasoners be able to keep track of all their arguments back to their foundational states, they still aim to approximate the foundationalism in Pollock’s account of ideal reasoning and require that reasoners keep track of the most recent step in the chain of reasoning leading to a particular belief.
immune to doubt in the form of questions about entitlement, but such
questions themselves stand in need of some sort of warrant or justification.
Entitlement is, to begin with, a social status that a performance or
commitment has within a community. (MIE 177)

Could we then use Brandom’s default entitlements as Pollock’s foundational
states? Commitments to which one has default entitlement can be understood as
epistemic starting places in the sense that they do not need to have been justified in order
to justify other beliefs. However, the set of claims that one has default entitlement to is
itself quite flexible and can change over the course of a conversation. Brandom writes,

There is no fixed point in advance where demands for justification or
demonstration of entitlement come to an end, but there are enough places
where such demands can end that there need be no global threat of
debilitating regress. (MIE 178 emphasis in original).

On Pollock’s account we are able to resolve conflicting reasons by tracing chains
of justification back to a predetermined and static epistemic starting point. Once we
remove the epistemic starting point, all we can make of the idea that someone has an
ultimately undefeated argument for commitment p is that that person does not have any
beliefs (commitments) that are incompatible with either p or the entitlement-preserving
inference leading to p. This (in some sense) is already Brandom’s position, and therefore
Pollock’s account wouldn’t offer anything.

---

41 I am here treating ‘justified’ as an act rather than as a normative status. There is no
contradiction in saying that one is entitled to commitment p, but one’s commitment to p
has not been justified.
The more serious difficulty we would encounter if we tried to incorporate Pollock’s account of defeasible reasoning into Brandom’s general theory of reasoning is that reasoning on Pollock’s account is individualistic, while on Brandom’s account reasoning is necessarily social. For Pollock the distinction we make between someone’s reasoning correctly and someone’s reasoning incorrectly seems to lie in that person’s successfully or unsuccessfully following all the rules of reasoning. Brandom, on the other hand, is not interested in identifying some standard which we can use to test whether someone is reasoning correctly or not. Brandom is more interested in how different reasoners (as scorekeepers) each draw different distinctions between correct and incorrect reasoning—how each reasoner distinguishes what someone is really committed and entitled to (according to that scorekeeper) and what someone merely takes himself or herself to be committed and entitled to (according to the same scorekeeper). This is one of the most central and interesting features of Brandom’s general theory of language and reasoning, and it is unclear how Pollock’s account of defeasible reasoning could be modified to accommodate it.

One of the features of Pollock’s more individualistic account that would be difficult to incorporate into Brandom’s more social account is Pollock’s appealing to beliefs in conditionals. For example if scorekeeper S1 keeps score on scorekeeper S2, S1 would need to keep track of whether or not S2 believes \(~(p \rightarrow q)\) when assessing

---

42 I will set aside the discussion of whether Pollock holds what Brandom calls ‘regulism’. In “Defeasible Reasoning” I don’t think Pollock is particularly interested in explaining how reasoners themselves distinguish good reasoning from bad reasoning. He may be better understood as positing “rules of reasoning” in an attempt to explain how people in fact reason—i.e. as doing a sort of psychology.
whether S2’s prima facie argument from \( p \) to \( q \) is undercut or not.\(^4\) The problem is that Brandom’s scorekeeping account of primitive linguistic practices (using languages without logical vocabulary) cannot appeal to beliefs in conditionals because it is supposed to explain how language users’ practices institute the implicit inferential proprieties that they can make explicit in the form of conditionals once they have logical vocabulary. Simply put, by allowing reasoners to simply list conditionals in their set of beliefs, Pollock just \textit{assumes} just what Brandom is trying to explain: what one must do in order to have inferential commitments.

Brandom \textit{does} allow for we logically competent theorists to interpret primitive language users as having \textit{implicit} commitments to inferential relations (which we might express using conditionals or other logical vocabulary). But Brandom \textit{does not} allow us to interpret primitive language users as attributing inferential commitments to other primitive language users because primitive language users cannot themselves understand the contents of such commitments. “Sentential logical vocabulary, paradigmatically the conditional, makes it possible to attribute acknowledgement of specific \textit{inferential} commitments.” (MIE 640, italics in original).

The same argument applies to the way that Pollock’s account allows defeaters that are not refuters. On Pollock’s account, a reasoner can be committed to a defeated inferential relation in virtue of their having it represented in their set of beliefs (perhaps as a conditional). But, again, the primitive language users Brandom is interested in describing do not have the vocabulary necessary for holding such beliefs.

\(^4\) See Pollock’s belief update rules 512-514.
There are various other complications that I will not examine in detail here, such as Brandom’s claiming that scorekeepers keep score according to their own inferential commitments rather than those of the person on whom they keep score. All of the issues that I have raised here are, I think, rather serious challenges to overcome.

**PART 3: Raymond Reiter’s Default Logic**

**Presentation of the Position:**

I will now explore Raymond Reiter’s “A Logic for Default Reasoning” as a representative of an account of defeasible reasoning from the AI literature that we might be able to integrate into Brandom’s theory of reasoning. Default logic is a good representative because it is one of the more common approaches to understanding defeasible reasoning in AI and because default rules act similarly to Brandom’s entitlement-preserving relations. Most of the more technical details of default logic are not necessary for the points being made here and therefore I will be presenting the position quite informally.

Reiter, and I think most of the AI community, see defeasible reasoning (which includes default reasoning) as letting us fill in some of the gaps in our knowledge that could not be filled in by more traditional deductive reasoning. Default reasoning

---

44 See also McCarthy 1987.
corresponds to the process of deriving conclusions based on patterns of
inference of the form ‘in the absence of any information to the contrary,
assume…’. Reasoning patterns of this kind represent a form of plausible
inference and are typically required whenever conclusions must be drawn
despite the absence of total knowledge of the world (Reiter 68)

Using the most common example in the AI literature, we want to be able to infer
“Tweety can fly” from “Tweety is a bird”, in the absence of knowledge to the contrary
(such as Tweety’s being a penguin, an ostrich, or dead). We don’t want to have to list all
the exceptions to the rule $IF \ Bird(x) \ THEN \ Flies(x)$, so it enough to say unless we
have information to the contrary we assume that if something is a bird then it flies. Reiter
represents this default rule as follows:

$$\begin{align*}
\text{BIRD}(x) \land \text{FLY}(x) \\
\text{FLY}(x)
\end{align*}$$

The $M$ above should be understood as “It is consistent to assume”, allowing us to
interpret the default rule as saying “If x is a bird and it is consistent to believe that x can
fly then one may believe that x can fly” (Reiter p. 69). It is not necessary that the
expression immediately following $M$ and the consequent of the inference be identical;
sometimes we may wish to include additional claims that must be consistent with one’s
belief set before one may believe the consequent.
One of the interesting features of Reiter’s defaults is that they are highly permissive in that they would permit someone to update their beliefs in a variety of different ways depending upon which defaults the reasoner applies to their belief set first. Consider the following example given by Reiter (p.70)

Default 1  “Assume a person’s hometown is that of his/her spouse.”

\[
SPOUSE(x, y) \& hometown(y) = z : M \text{ hometown}(x) = z \\
hometown(x) = z
\]

Default 2  “Assume a person’s hometown is where his/her employer is located.”

\[
EMPLOYER(x, y) \& location(y) = z : M \text{ hometown}(x) = z \\
hometown(x) = z
\]

Let us suppose that \(x\) is Mary and that we know that Mary’s spouse is located in Toronto and Mary’s employer is located in Vancouver. With these facts and with those two defaults there are two different ways that we can extend our knowledge—giving us two permissible sets of beliefs. If we apply the first default before the second then according to the first default we can assume that Mary’s hometown is Toronto, but cannot subsequently use the second default to assume that Mary’s hometown is Vancouver (because our believing Mary’s hometown is Toronto is inconsistent with believing Mary’s hometown is Vancouver). However, if we apply the second default to our set of beliefs before the first, the situation is reversed such that we can believe Mary’s hometown is Vancouver, but not that Mary’s hometown is Toronto. So depending upon how we apply the given defaults to our original set of beliefs we can end up with one of two different sets of expanded beliefs.
Reiter’s default reasoning therefore allows multiple extensions, where the ‘extension’ of a set of defaults $D$ is the set of beliefs resulting from applying every default rule in $D$ that we can.

Defaults therefore function like meta-rules; they are instructions about how to create an extension of this incomplete theory [about the world]. Those formulae sanctioned by the defaults and which extend the theory can then be viewed as beliefs about the world. Now in general there are many different ways of extending an incomplete theory, which suggests that the default rules may be nondeterministic. Different applications of the defaults yield different extensions and hence different sets of beliefs about the world. (Reiter 71)

There are some clear similarities between Reiter’s defaults and Brandom’s entitlement-preserving relations. Brandom would take “IF Tweety is a bird THEN Tweety can fly” to express an entitlement-preserving relation such that persons who are committed and entitled to ‘Tweety is a bird’ are consequently entitled to ‘Tweety can fly’ but only if that person is not yet committed to anything incompatible with ‘Tweety can fly’. Using the same example, Reiter likewise claims that persons who believe ‘Tweety is a bird’ are only permitted to believe ‘Tweety can fly’ if they do not believe anything inconsistent with ‘Tweety can fly’.

Another similarity lies in how Brandom’s entitlement-preserving relations are permissive and nondeterministic much like Reiter’s sets of defaults are. If we turned the
defaults in Reiter’s hometown example into Brandom’s entitlement-preserving relations, we would end up with largely the same conclusions that Reiter does. Suppose we are committed and entitled to “Mary’s spouse lives in Toronto” and “Mary’s employment is in Vancouver”, and are also committed to the following inferential relations:

“Mary’s spouse lives in Toronto” > “Mary’s hometown is Toronto”

“Mary’s employment is in Vancouver” > “Mary’s hometown is Vancouver”

“Mary’s hometown is Toronto” // “Mary’s hometown is Vancouver”

On Brandom’s account we are entitled to undertake commitment to either ‘Mary’s hometown is Toronto’ or ‘Mary’s hometown is Vancouver’ so long as we have not yet actually undertaken commitment to either. In light of these inferential relations and the original set of commitments we acknowledge, there are (at least) two different ways that we are entitled to expand upon our set of commitments. This roughly corresponds to Reiter’s claiming that there are two different extensions satisfying the set of facts $W$ and the set of defaults $D$ in his example.\(^\text{45}\)

Difficulties Incorporating Features of the Position

\(^{45}\) I say it only roughly corresponds because, as will be explained shortly, on Brandom’s account someone’s changing beliefs results in that person’s changing their inferential commitments.
There is no doubt that Reiter’s default logic is nonmonotonic, but given that Reiter’s defaults are so similar to Brandom’s entitlement-preserving relations, it is unlikely that Reiter’s account would allow entitlement recovery. But rather than argue that default reasoning cannot allow it, I will instead argue that even if it (or one of the later elaborations on it) did allow entitlement recovery, there would be serious difficulties incorporating it into Brandom’s account of reasoning—despite Reiter’s defaults being so similar to Brandom’s entitlement-preserving relations.

The first difficulty we would face if we were to incorporate default logic into Brandom’s account arises from the fact that default logic is a system of logic rather than an account of belief updating (or scorekeeping). Default logic will tell us that we can infer particular conclusions from a given set of beliefs (acting as premises) and a given set of defaults, but it does not tell us how we should change our beliefs in light of new information—particularly if the new information is a reason against one of our previously held beliefs. If we consider the match-lighting example, we can use default logic to explain why someone who comes to believe Struck should believe Lights. However, at the next stage of the example we cannot use default logic to explain why the reasoner who now believes Struck, and Lights, should cease accepting Lights in response to her coming to believe EMF. It may be tempting to say that the reasoner should cease accepting Lights because it is no longer consistent with the rest of her beliefs (which was one of the conditions necessary for inferring Lights in the first place). However, the consistency condition (expressed in defaults using $M$) only applies when inferring a conclusion from premises, not for maintaining consistency within the set of premises (of which Lights is now a member).
Another challenge that we would face if we tried to incorporate default logic into Brandom’s account is that the set of defaults to be used in a reasoning scenario is largely independent of what the facts are (or rather, the defaults that a reasoner accepts is unconstrained by what that reasoner believes). We saw in the previous chapter how Brandom is committed to an inferential holism where a reasoner’s inferential commitments largely depend upon the assertional commitments the reasoner acknowledges. One consequence of this is that a reasoner may implicitly revise her inferential commitments in light of her taking herself to have gained or lost assertional commitments. This quite clearly plays no role in Reiter’s default logic, where a reasoner’s applying a default to her belief set may result in the reasoner’s acquiring a new belief, but not in the reasoner’s gaining or losing a commitment to a default. Simply put, default logic does indeed have defeaters which are not refuters, but only because it does not allow for refuters. This is not acceptable, because (as explained in Chapter III) even though Brandom’s account requires defeaters that are not refuters, it also requires that claims and inferential relations can be refuted.

The inability to accommodate inferential holism is by no means unique to Reiter’s account. I think that it is fair to say that most accounts of defeasible reasoning in the AI literature are happy to either stipulate a static set of inferential relations, or attempt to give some account of all and only those conditions in which someone could rationally accept different inferential relations.\[^{46}\] While these might be acceptable moves to make in the AI literature, they are incompatible with what Brandom is trying to do and suggest

\[^{46}\] This is also true of McCarthy’s Circumscription approach to defeasible reasoning (see McCarthy 1987). As we shall later see, the AGM account does not assume static rules and actually accommodates holism rather well.
that Brandom’s project is importantly different from the project undertaken by most AI theorists.

Another distinctive feature of Brandom’s broader theory that makes it difficult to incorporate other accounts of defeasible reasoning is the explanatory priority of the propositional, which effectively reverses the order of explanation adopted in most other accounts of defeasible reasoning. Consider the example of a default that Reiter gives:

\[
BIRD(x) : \text{MFLY}(x) \\
FLY(x)
\]

By substituting ‘Tweety’ for the variable \(x\), we can explain defeasible inferences involving whole sentences (such as ‘Tweety can fly) in terms of general patterns of inference involving only predicates (such as ‘\(x\) can fly’ or \(Fly(x)\)). This (admittedly very plausible) approach is adopted in many if not most accounts of defeasible reasoning.\(^{47}\) One of the benefits of doing so is that it leaves room open for a reasoner to accept the general pattern of inference \(Bird(x) > Flies(x)\) while denying the particular inference \(Bird(\text{Tweety}) > Flies(\text{Tweety})\).\(^{48}\)

Brandom, however, cannot adopt this approach because for him the basic units of meaning are whole sentences, as opposed to individual words or phrases expressing predicates or singular terms. The reason is simply that only whole sentences can stand in inferential relations—can be used to give or ask for reasons. It is only once we have

\(^{47}\) Another good example of this can be seen in John McCarthy’s “Circumscription—A Form of Non-Monotonic Reasoning”
\(^{48}\) This seems to be the approach that Pollock adopts.
identified the meaning (as inferential roles) of various sentences that we can then identify the meanings of sub-sentential expressions in terms of their contribution to inferential roles of different sentences.

Since semantics must in this way answer to pragmatics, the category of sentences has a certain kind of explanatory priority over subsentential categories of expressions, such as singular terms and predicates. For sentences are the kind of expression whose freestanding utterance (that is, whose utterance unembedded in the utterance of some larger expressions containing it) has the pragmatic significance of performing a speech act…[Subsentential] expressions cannot have semantic contents in the same sense in which sentences can. They can be taken to be semantically contentful only in a derivative sense, insofar as their occurrence as components in sentences contributes to the contents (in the basic, practice-relevant sense) of those sentences. (MIE 363-364)

Once someone grasps the conceptual content of various subsentential expressions then they will be able to understand and use new sentences composed from these expressions, but all of this assumes a prior competence at using whole sentences in one’s giving and asking for reasons (MIE 364-365). For present purposes we need not concern ourselves with the finer details. What matters is that on Brandom’s account one must have many inferential commitments relating a lot of different sentences in order to have a view on the conceptual contents of predicates and singular terms. As a result Brandom
must be able to give an account of what it means for someone to be committed to
defeasible inferential relations like *Tweety is a bird > Tweety can fly* without
appealing to general patterns of inference involving predicates like *Bird(x) > Flies(x)*.

**PART 4: AGM on Belief Revision and Nonmonotonic Inference:**

**AGM and Belief Revision**

The last theory of reasoning that I will be considering is often simply referred to
as AGM. AGM is an account of belief updating (or theory change) first developed by
three authors: Carlos Alchourron, Peter Gärdenfors, and David Mackinson. While these
authors usually talk about *theory* change, they also talk about it in some contexts as *belief*
updating. Both consist of rationally adding and removing propositions from a set of
propositions. I will talk about it just in terms of *belief* updating.49 Today AGM is
probably best understood as a research program that originated with the three mentioned
authors but is carried on by many other theorists.50 I will be focusing primarily on the
original account and a few of the developments by the original authors.

49 I will be using ‘update’ to refer to the general practice of rationally changing one’s
beliefs (or score), treating each of the operations on a belief set discussed in the AGM
account (revision, contraction and expansion) as falling under belief updating generally.
I recognize that ‘update’ is sometimes used to refer to another specific type of operation
that could be performed on one’s belief; this is not how I am using it here.
50 For example see: Katsuno and Mendelzon (1992), Hansson et al. (2001), and Eduardo
and Rott (2004).
What makes AGM especially relevant to this dissertation is that it provides an account of belief revision which two of the three authors later use to develop an account of non-monotonic inference. The authors, like Brandom, appeal to belief updating to explain what is inferentially related to what else. Not only that, but AGM is compatible with inferential holism and the explanatory priority of the propositional.

One can think of the core of the original AGM account as consisting of two different parts: one is a set of postulates for belief revision (put forth originally by Gärdenfors). These postulates do not indicate how beliefs should be updated, but they purport to specify some minimal conditions for successful expansion, contraction, and revision of beliefs. I will not be focusing on these postulates because they assume that beliefs will be updated according to a monotonic consequence operator. Instead I will focus on AGM’s more constructive account of belief contraction and revision which the AGM proponents take to satisfy those postulates.\textsuperscript{51}

There are three different operations that a reasoner can use to update their theory (their set of beliefs) \( A \) by some proposition \( x \). Each operation is performed on a set of propositions (or beliefs) and a particular proposition, and yields a logically closed updated set of beliefs.

Expansion: \( A + x \)

Contraction: \( A - x \)

Revision: \( A * x \)

\textsuperscript{51} This is primarily drawn from the three authors’ paper *On the Logic of Theory Change: Partial Meet Contraction and Revision Functions*. (1985)
Both expansion and revision consist of adding a new belief $x$ to one set of accepted beliefs $A$. If $x$ is consistent with $A$ then expansion *can* be performed (but so can revision). If $x$ is inconsistent with $A$ then revision *must be* performed. Expansion is the most straightforward and simply consists of adding $x$ to $A$ and then all the classical consequences of $A \cup x$. Contraction consists of removing a belief $x$ from set of beliefs $A$ and then removing various further beliefs such that the resulting set of beliefs will not have $x$ as a logical consequence. To put the matter another way, someone’s contracting $A$ by $x$ consists of that person no longer accepting set of propositions $A$ and instead accepting some logically closed subset of $A$ which does not include $x$. Contraction is much more complicated than Expansion because logic alone does not determine which subset of $A$ the reasoner should accept.

Revision faces the same complications as contraction because if a reasoner revises her set of beliefs $A$ by $x$ and $x$ is incompatible with $A$ then various beliefs must be removed from $A$ in order to make it consistent with $x$. But, once again, logic alone does not determine which particular set of beliefs should be removed from one’s set of beliefs in order to make it consistent with $x$.

Because expansion is generally well understood, the focus of the AGM model is on contraction and revision. Revision is understood in terms of contraction. Treating ‘$Cn$’ as a classical logical consequence operation, $A * x = Cn((A - \neg x) \cup \{x\})$. $A - \neg x$ is the set of beliefs that results from contracting $\neg x$ from $A$, thereby yielding a set of beliefs $A'$ that is consistent with $x$. This gives us $A * x = Cn(A' \cup x)$, where $Cn(A' \cup x)$ is the logical closure of the union of $A'$ and $x$. 
The challenge to be overcome in providing an account of contraction can be seen in the following simple example. Suppose $A = \{a, b, \sim a \lor b\}$ and we then want to contract $A$ by $b$. Assuming that sets of beliefs are consistent and logically closed it follows that the contraction of $A$ by $b$ cannot merely consist of removing $b$ from $A$ and therefore we must also remove some propositions from $A$ such that $b$ is no longer a consequence. Often, however, there is no unique way of changing a set of beliefs so that it will no longer imply some proposition. In order to make it so that $b$ is no longer a consequence of $A$ in the present example, we can remove $a$ or $\sim a \lor b$, or both. But logic alone does not indicate which of these options should be pursued.

The AGM model of belief contraction and revision attempts to uniquely specify which beliefs should be removed and which should be kept when contracting a proposition from a set of beliefs. One of the central aims of the AGM account of contraction, called *partial meet contraction*, is that change in belief be minimal and that reasoners ought to reject as little as they can when contracting a belief. This is why the authors appeal to *maximal* subsets of the set of beliefs being contracted, using the notation $A \downarrow x$ to mean the set of all maximal subsets $B$ of $A$ such that $B$ does not imply $x$.

Of course $A \downarrow x$ picks out a set of maximal subsets rather than a unique one. One proposal considered by two of the authors for picking out a unique subset consists of using the intersection of $A \downarrow x$. The resulting contraction function where $A - x = \cap (A \downarrow x)$ is called *meet* (or *full meet*) contraction. The problem with full meet contraction (as the authors acknowledged) is that the resulting set of beliefs may end up far too small
in cases where there is very little overlap between some of the maximal subsets that do not imply $x$. \footnote{See C Alchourron and D Mackinson \textit{On the Logic of Theory Change: Contraction Functions and their Associated Revision Functions} (1982)}

It was largely in response to this problem that the \textit{partial} meet contraction function was adopted. On this account the set of maximal subsets considered is reduced to only those that are selected as the most epistemically important by some selection function $y$. Partial meet contraction functions can therefore be defined as $A - x = \cap y(A \models x)$. The hope is that the selection function $y$ will limit the maximal sets of beliefs considered so that their intersection will not be too small.

Naturally, if there is a selection function that picks out those maximal sets of propositions that are most epistemically important then there must be some ordering or ranking amongst maximal sets of beliefs. One of the original AGM authors, Gärdenfors, attempts to explain what it means for one proposition to be more epistemically important than another in terms of how entrenched the proposition is —how difficult it is to contract. \footnote{See P Gärdenfors \textit{Epistemic Importance and Minimal Changes of Belief}} \footnote{In that paper he talks of ordering sentences rather than propositions, but given that partial meet contraction and revision functions applies to propositions I assume that he takes the discussion in \textit{Epistemic Importance} to apply to propositions as well.} Gärdenfors then goes on to explain epistemic entrenchment in terms of belief contraction. We can know whether $p$ is at least as entrenched as $q$, which he expresses as $p \geq q$, if $q$ remains in a set of beliefs after contracting $p \& q$ from it. Let $K$ be the initial set of beliefs. If $q$ is not a member of $K - (p \& q)$ then $q$ cannot be more epistemically entrenched than $p$. 

\begin{itemize}
\item It was largely in response to this problem that the \textit{partial} meet contraction function was adopted. On this account the set of maximal subsets considered is reduced to only those that are selected as the most epistemically important by some selection function $y$. Partial meet contraction functions can therefore be defined as $A - x = \cap y(A \models x)$. The hope is that the selection function $y$ will limit the maximal sets of beliefs considered so that their intersection will not be too small.
\item Naturally, if there is a selection function that picks out those maximal sets of propositions that are most epistemically important then there must be some ordering or ranking amongst maximal sets of beliefs. One of the original AGM authors, Gärdenfors, attempts to explain what it means for one proposition to be more epistemically important than another in terms of how entrenched the proposition is —how difficult it is to contract. Gärdenfors then goes on to explain epistemic entrenchment in terms of belief contraction. We can know whether $p$ is at least as entrenched as $q$, which he expresses as $p \geq q$, if $q$ remains in a set of beliefs after contracting $p \& q$ from it.
\item If $q$ is not a member of $K - (p \& q)$ then $q$ cannot be more epistemically entrenched than $p$.
\end{itemize}
This same approach is used to rank the epistemic importance of maximal subsets of beliefs (which do not contain the proposition being contracted). Because the AGM account assumes that the belief sets being considered are logically closed, it follows that each subset \( K' \) of a belief set \( K \) will be represented in \( K \) as a conjunction of sentences we can call \( S_{K'} \). We can then assess whether \( K' \geq K'' \) on the basis of whether \( S_{K'} \) is a member of \( K \setminus (S_{K'} \& S_{K''}) \).

Gärdenfors however, admits that the problem with explaining epistemic importance in this way is that he ends up in an explanatory circle. He appeals to epistemic importance in order to explains contraction, and explain contraction by appealing to epistemic importance. Beyond the somewhat circular explanation of epistemic importance, Gärdenfors speculates that epistemic importance is likely determined by more than the set of beliefs that we hold and that pragmatic considerations must play an important role.

I suspect that theoretical considerations will in many cases not give any guidance as to which of two sentences is the epistemically more important. I believe that in these cases one has to rely on pragmatic factors. For example, when a person contracts his state of belief in a debate or when giving an explanation, the conversational context may be a crucial factor in determining the current epistemic importance of the sentences under consideration.\(^{55}\)

\(^{55}\) Gärdenfors, *Epistemic Importance and Minimal Changes of Belief*, 151, emphasis in original.
Gärdenfors makes similar remarks about the impact that a change of belief can have upon the ordering of epistemic importance. He says that if counterfactual conditionals were amongst reasoners’ beliefs then we could look to these to help identify what ordering of epistemic importance someone would accept were they to have different beliefs. But Gärdenfors goes on to say, “This operational method gives us, however, little understanding of the underlying factors that determine these changes of epistemic importance.”

AGM and Nonmonotonic Inference

Makinson and Gärdenfors argue that belief revision and nonmonotonic reasoning are in fact two sides of the same coin. To start, for any set of propositions \( A \) and any element of \( A \), we take \( A \vdash x \) to hold, whereby \( \vdash \) is a nonmonotonic inference relation. Next, and this is the crucial move, we treat the revision function \( A \ast x \) to be an operation on \( x \) involving \( A \), rather than as an operation on \( A \) (keeping in mind that \( x \) will always be a member of \( A \ast x \)). In light of this we get \( x \vdash y \) iff \( y \in A \ast x \). The set of propositions \( A \) therefore functions as a background theory (or set of collateral premises) relative to which a proposition \( x \) will stand in different nonmonotonic inference relations.

To see that \( \vdash \) is in fact nonmonotonic we can consider the following example. Suppose that \( A = Cn(\sim a \lor \sim b) \). In light of this \( A \ast (\{a \& b\}) \) will be some consistent set containing \( a \& b \), and \( A \ast (\{a\}) \) will be some consistent set containing \( a \& \sim b \) (which

56 Ibid 152
57 Makinson and P Gärdenfors Relations Between the Logic of Theory Change and Nonmonotonic Logic
is the classical consequence of adding proposition $a$ to $\sim a \lor \sim b$.) Because $\sim b$ is a member of $A \ast \{\{a\}\}$ it follows that $\{a\} \vdash \{\sim b\}$. Furthermore, because $\sim b$ is not a member of $A \ast \{(a \& b)\}$, we do not get $\{a, b\} \vdash \{\sim b\}$. And because $\{a, b\}$ is a superset of the set of $\{a\}$ it follows that the inference relation $\vdash$ is nonmonotonic. Makinson and Gärdenfors write,

With this gestalt, we are in a position to understand the link between nonmonotonic logic and the logic of theory change. The key idea is:

1. See the revision of a theory $A$ by a proposition $x$, forming a theory $A \ast x$, as a form of nonmonotonic inference from $x$;

2. Conversely, see a nonmonotonic inference of a proposition $y$ from a proposition $x$ as a discovery that $y$ is contained in the result of revising a fixed background theory $A$ so as to integrate $x$. In this way, the nonmonotonic relation $x \vdash y$ serves as a shorthand for $x \vdash_A y$ which indicates that the nonmonotonic inference is dependent on the background theory $A$. (Makinson and Gärdenfors 188-189).
Difficulties Incorporating Features of the Position

As we begin to assess what AGM may be able to contribute to Brandom’s account of defeasible reasoning, I will start by acknowledging that AGM’s nonmonotonic inference relation allows entitlement recovery. This can be seen in the following example. Suppose that \( A = \{\neg a \lor d, \neg (a \& b) \lor \neg d, \neg c \lor d\} \) and these propositions have the following ranking in epistemic entrenchment \( (\neg c \lor d) > (\neg (a \& b) \lor \neg d) > (\neg a \lor d) \). It would plausibly follow from this that

\[
d \text{ is a member of } A \ast (\{a\}) \quad \text{and therefore} \quad \{a\} \vdash \{d\}
\]

\[
d \text{ is not a member of } A \ast (\{a \& b\}) \quad \text{and therefore} \quad \text{Not } \{a, b\} \vdash \{d\}
\]

\[
d \text{ is a member of } A \ast (\{a \& b \& c\}) \quad \text{and therefore} \quad \{a, b, c\} \vdash \{d\}
\]

Furthermore, many of the positions adopted by AGM (particularly by Gärdenfors) in regards to the relation between belief updating and inference/logic overlap in important ways with Brandom’s account of propositional content and the circumstances for attributing inferential commitments.\textsuperscript{58} Not only does AGM try to explain nonmonotonic

\textsuperscript{58} Ten years before Brandom wrote MIE, Gärdenfors introduces his paper “The Dynamics of Belief as a Basis for Logic” as follows,

The aim of this article is to give a reconstruction of propositional logic from an epistemological basis. The fundamental notions will be states of belief and changes of belief. A central idea in the reconstruction is that propositions are
inferential relations in terms of belief updating (like Brandom does), it is also holistic in the sense that what nonmonotonically follows from a claim will depend upon a background set of collateral premises.

There are two different ways that I think that we could plausibly try to integrate features from AGM into Brandom’s broader account of reasoning. First, we might try to incorporate aspects of AGM’s account of belief updating into Brandom’s account of the consequences of attributing inferential commitments (so that Brandom’s account would allow entitlement-recovery). Second, we might try to incorporate aspects of AGM’s account of nonmonotonic consequence into Brandom’s account of the circumstances for attributing inferential commitments (so that Brandom’s account would allow defeaters that are not refuters). I will now discuss the first option.

The first difficulty that we would encounter if we tried to incorporate AGM’s account of belief updating into Brandom’s is that the AGM account of belief updating does not actually tell us anything about defeasible inferential relations. The AGM account might show that belief revision is, in a sense, nonmonotonic, but the account of revision itself cannot be usefully incorporated into Brandom’s broader account because it makes no appeal to defeasible inferential relations, and assumes classical logical inference. This is perhaps most clearly seen in the fact that an inferential relation’s being defeated simply plays no role in the AGM account. As a result, the AGM’s account of belief updating

identified with a certain kind of changes of belief. On this interpretation, a proposition is characterised by the change it would induce if added to a state of belief. Formally, propositions will be defined as functions with states of belief as arguments and values. These functions will be assumed to satisfy certain requirements connected with how the addition of the information of a proposition would affect the content of a state of belief.
does not actually tell us anything about the consequences of one’s having commitments to defeasible inferential relations, nor does it deserve to be called an account of reasoning that is defeasible (in the inconclusive sense of ‘defeasible’). As far as I can tell, the original authors don’t describe it as such either.

Even if we bracket this difficulty, we might still want to consider whether AGM’s epistemic entrenchment ranking could be usefully incorporated in Brandom’s account of how score should be updated when someone has competing reasons. Taking this approach, one might suggest that we modify Brandom’s account so that scorekeepers would accept an epistemic entrenchment ranking in addition to their accepting a set of claims and a set of inferential relations. There are, of course, many different ways that we might try to incorporate entrenchment rankings into Brandom’s account. But, as I will now argue, there are good reasons to doubt that this would be a more promising approach to the problem of entitlement-recovery than simply modifying the various consequences of one’s being committed to the three different types of inferential relation.

There is an important sense in which AGM’s epistemic entrenchment ranking plays much the role that Brandom’s inferential relations do. We assess what epistemic entrenchment rankings and inferential relations a reasoner accepts by looking at how that reasoner would update her beliefs (or score). Furthermore, what particular ranking and what particular inferential relations a reasoner accepts will depend on and vary with the set of beliefs that the reasoner happens to accept at the time in question. Once we have identified the epistemic ranking or inferential relations that a reasoner accepts (relative to her particular set of beliefs) we can then determine how that reasoner should revise her
beliefs upon learning new information—using AGM’s partial meet account of revision, and Brandom’s scorekeeping algorithm.

The strong functional similarity between entrenchment rankings and inferential relations suggests that Brandom’s account should not need entrenchment rankings because it already has inferential relations on board (since Brandom is, after all, providing an *inferentialist* theory of meaning). It is not even clear how AGM’s account of epistemic entrenchment would work in Brandom’s broader account. For instance, it is not clear how ranking a scorekeeper’s assertional commitments would help overcome the problem of entitlement-recovery if scorekeepers update scores according to their inferential commitments.

It might be a little more obvious how some ranking of inferential commitments (as opposed to assertional commitments) might help, but then it is unclear how we could use the AGM account to identify such a ranking: we would have to somehow figure out which of two inferential relations a scorekeeper would remain committed to if she were to contract the conjunction of them from her set of inferential commitments. First of all, a conjunction of inferential commitments does not make a whole lot of sense, unless we treat inferential commitments as assertional commitments. Second of all, Brandom’s scorekeepers do not update their inferential commitments in the relevant sense; we would first need an account of the circumstances for attributing inferential commitments.

I will now argue that it is unlikely that we can usefully incorporate AGM’s account of nonmonotonic inference into Brandom’s account of the circumstances for attributing inferential commitments (in an attempt to overcome the problem of defeaters...
that are not refuters). What seems to be the most plausible way to incorporate AGM’s nonmonotonic consequence relation is to treat it as a possible account of the circumstances for attributing inferential commitments—particularly to entitlement-preserving inferential relations. Assuming that \( A \) is the background set of collateral premises then \( p > q \) if \( q \in A \ast p \). Alternatively, we could replace the revision function with the ScoreUpdate function, giving us

\[
S \text{ is implicitly committed to } p > q \text{ if } \\
E(q) \in \text{ScoreUpdate}(S, A, +CE\{p\}) \\
A = \text{Whatever score set } S \text{ takes herself to have}
\]

There are two serious problems that arise from using this as Brandom’s account of the circumstances for attributing commitments to entitlement-preserving relations. First of all, it would make any claim that was already a member of \( A \) count as an entitlement-preserving consequence of \( p \). \( S \) might not take the addition of commitment and entitlement \( p \) to score set \( A \) to have any impact on the presence or absence of entitlement \( q \), but on the proposed account we would still interpret \( S \) as being committed to \( p > q \) if \( E(q) \) were already a member of \( A \). Brandom’s original account is able to avoid this problem by focusing only on the difference that the addition or removal of commitments and entitlements makes to a score set. This is what the \( ScoreDiff \) function captured in the previous chapter.

The second problem with the proposed account of the circumstances for attributing commitments to entitlement-preserving relations is that it fares little better than Brandom’s original position at allowing defeaters that are not refuters. This is
because on Brandom’s original account, and on the modified AMG account, the
reasoner’s present set of inferential commitments are dependent upon the reasoner’s
present set of beliefs/commitments acting as the background set of collateral premises.

Assuming that $E\{q\}$ is not already a member of $A$, it follows that if the scorekeeper takes
herself to be committed to a defeater for $p \rightarrow q$ then $E\{q\}$ will not be a member of
$ScoreUpdate(S, A, +CE\{p\})$. As a result, we would not be entitled to attribute
commitment $p \rightarrow q$ to the scorekeeper, but would be entitled to interpret the scorekeeper
as taking the defeater to be a refuter.

Before proceeding, I am going to identify one final complication that we would
likely encounter if tried to incorporate aspects of AGM, or almost any other theory about
defeasible reasoning, into Brandom’s account of reasoning. Most accounts of reasoning
are only concerned with how reasoners update beliefs, roughly corresponding to
commitments in Brandom’s account. But reasoners (as scorekeepers) in Brandom’s
account don’t just update sets of commitments (or beliefs); they also update sets of
entitlements. It is core to Brandom’s account that one can be committed to claims that
one is not entitled to, and one can be entitled to claims that one is not committed to.
Therefore, entitlements must be treated differently than commitments. This adds a whole
other layer of complexity to the obstacles identified above (recognizing of course that the
amount of complexity added will vary from case to case).
Conclusion:

In this chapter I have presented three well established theories about defeasible reasoning (understanding ‘defeasible’ in a rather broad sense). I have also presented a number of obstacles that we would likely encounter if we tried to incorporate them (or features of them) into Brandom’s broader account of reasoning. My hope is that the various obstacles that I have identified are representative of the sorts of obstacles that we would likely encounter if we tried to incorporate most other theories of defeasible reasoning—other than the three specific ones considered here. Again, I am not arguing that these challenges (or numerous others not considered here) could not be overcome in some way or other. I am simply arguing that the mainstream literature on defeasible reasoning is unlikely to reveal any clear and simple way to fix Brandom’s account so that it would allow entitlement-recovery and allow defeaters that are not refuters. I hope, at this point, to have convinced the reader that the suggestion to incorporate other accounts of defeasible reasoning into Brandom’s would likely introduce more obstacles than it would overcome.

However, before moving on, I would like to acknowledge the usefulness of one idea that is commonly appealed to (and sometimes debated) in the AI literature, called the principle of specificity.\(^59\) The idea, very roughly, is that if one accepts the premises of two competing defeasible inferential relations (or rules) then one should give priority to inferential relations whose premises draw from more knowledge rather than less knowledge. In Chapter V, I will adopt a version of this principle to solve the problem of entitlement-recovery.

\(^{59}\) For Example, See Poole (1985)
CHAPTER V:

Solving the Problem of Entitlement-Recovery:

The Revised Account of the Consequences of Attributing Commitment to the
Three Different Types of Inferential Relations

Reviewing the Problem

In this and the next chapter I will propose modifications to Brandom’s account of reasoning that will overcome the two problems that I have argued it faces. The first of these problems, discussed in Chapter II, was that Brandom’s account did not allow entitlement-recovery. As we saw, we could not use Brandom’s three types of inferential relations to interpret a scorekeeper as obliged to update score correctly in Brandom’s own match-lighting example. This indicated that there was a problem with the consequences of the three types of inferential relations.

The second of the problems was discussed in Chapter III, this being the problem of defeaters that are not refuters. The problem was that the circumstances for attributing an inferential commitment to a scorekeeper would no longer be satisfied on any occasion on which that scorekeeper took that inferential relation to be defeated for herself. As a result she would not be obliged to update the scores of her interlocutors according to that inferential relation even if the inferential relation was not defeated for those interlocutors.
In order to modify Brandom’s three inferential relations so that his account would allow defeasible reasoning, we must therefore modify his accounts of the circumstances for and consequences of attributing inferential commitments. In this chapter I will propose the modifications that need to be made to the consequences.

**PART 1: Clarifying How the Expressions Commonly Employed in Discussions of Defeasible Reasoning will be Used.**

Before presenting my revisions to Brandom’s account I will clarify how I will be using a number of expressions commonly employed in discussions of defeasible reasoning. While in some sense *I could* avoid using many of these expressions, only appealing to the scorekeeping metavocabulary and the three types of inferential relations, doing so would make it difficult to motivate or evaluate the revised scorekeeping account of defeasible reasoning. We have little reason to interpret the adding and removing of commitments and entitlements as a sort of defeasible reasoning unless we can describe it using the vocabulary commonly employed in discussions of defeasible reasoning. The defeasible reasoning vocabulary will act as a bit of an entry point for discussing how we should understand the different types of inferential commitments. It will eventually be replaced by expressions formulated in purely Brandomian scorekeeping terms.

We know that we are supposed to understand material inferential relations as determining what is a reason or a defeater for what else, but there are a lot of relevantly
different senses in which we can understand something’s being a reason or a defeater for something else. I have already made some clarifications (particularly in Chapter I) in regards to defeaters and refuters, but I have not yet systematically presented how we should understand a lot of the related expressions (such as what it means for someone to have a reason, or a refuter). I will not be arguing that the way I will be using some of the expressions associated with defeasible reasoning is uncontroversial or is the only correct way to use them. My present aim is merely to articulate the sense in which we should understand these expressions within a Brandomian scorekeeping conception of reasoning.

To start, we need to recall from Chapter I that there are two different senses in which we might say that reasoning is defeasible.

3) The refutability sense: The acceptance or non-acceptance of particular beliefs and the rational relations between those beliefs.

Or

4) The inconclusive sense: Some property belonging to the rational relations between beliefs.

Throughout this dissertation I have been using ‘defeasibility’ (and its related expressions) in the inconclusive sense. If we simply want to say that someone may rationally reconsider their acceptance of a particular belief or rational relation, then we say that the belief or rational relation is refutable, rather than defeasible.
Defeasible relations of rational support are to be understood in terms of entitlement-preserving relations:

On Brandom’s account, relations of rational support take the form of material inferential relations (particularly commitment-preserving and entitlement-preserving relations). Because defeasibility is a feature of rational relations, a claim $p$ is a good defeasible reason for another claim $q$ if and only if there is some material inferential relation leading from $p$ to $q$ that has the property of being defeasible. Going on the assumption that entitlement-preserving relations are the only type of material inferential relations that are defeasible, it follows that $p$ is a good defeasible reason for $q$ if and only if $q$ is an entitlement-preserving consequence of $p$ (iff $p \rightarrow q$).

An inferential relation’s being materially good depends on score, while its being defeated does not:

Whether or not some entitlement-preserving relation $p \rightarrow q$ is materially good is independent of what commitments and entitlements anyone happens to have. In contrast, whether or not $p \rightarrow q$ is defeated will in part depend upon the particular set of commitments and entitlements being updated. Even though entitlement-preserving relations will be defeated when used to update the scores of some people but not defeated for other people, if an entitlement-preserving relation is materially good for one person

---

60 This claim is importantly different from the claim ‘Whether or not someone S takes $p \rightarrow q$ to be materially good is independent of what anyone (including S) takes themselves to be committed and entitled to.’ On Brandom’s account, whether or not an inferential relation is materially good depends on what the facts are, not on what anyone happens to be committed to. It is true that what we take to be the facts is identical with what we take ourselves to be committed to, but the inferential relations involving the claim ‘I am committed to p’ are importantly different from those involving just ‘p’ (or ‘p is a fact’).
then it is materially good for everyone (assuming that we are only discussing those that relate doxastic commitments and entitlements). This, on both my revised and Brandom’s original account, is necessary if we want to understand the material goodness of inferential relations as objective.

**Only inferential relations can have defeaters:**

Because defeasibility is a feature of some material inferential relations rather than of one’s commitments and entitlements, it is not possible for commitments or entitlements to claims to be defeated. It therefore follows that there are no defeaters for claims, but only for some types of material inferential relations—the entitlement-preserving ones.

**Defeaters are claims that defeat inferential relations:**

Very roughly, some claim $r$ is a defeater for a defeasible relation of rational support leading from $p$ to $q$ if and only if the entitlement-preserving relation $p > q$ is defeated whenever it is used to update the score of someone committed to $r$, or the score of someone with an interlocutor who has asserted $r$ and is entitled to $r$. This is exactly what would happen in Brandom’s (and my revised) account if $r$ stood in an incompatibility relation with $q$. (The specific conditions in which entitlement-preserving relations are defeated will be discussed in much more detail below.)

**Refuters are claims that are reasons against claims or inferential relations:**
When commitment to a claim \( r \) is a *reason against* one’s undertaking commitment to another claim \( q \), I will say that that \( r \) is a *refuter for* \( q \) rather than a *defeater for* \( q \). One way that \( r \) can be a refuter for \( q \) is by \( r \) and \( q \) being incompatible with each another. We even see this on Brandom’s original account, where one’s entitlement to \( q \) can be lost or prohibited by their undertaking commitment to \( r \). If \( p > q \) and \( r // q \) are materially good inferential relations then we would say that \( p \) is a *good defeasible reason for* \( q \), \( r \) is a *refuter for* \( q \), and \( r \) is a *defeater for* \( p > q \).61

Which claims are *defeaters* or *refuters* is independent of score; what claims are *defeated* or *refuted* depends upon score.

As with good reasons, we can determine whether some claim is a defeater or a refuter purely on the basis of the material inferential relations that hold between the various claims involved. The set of *defeaters* for a particular inferential relation are the same for everyone, so too for the set of *refuters*. Whether or not a claim has in fact been *refuted* or an inferential relation *defeated* will, however, vary from person to person and depend on whether or not he or she is committed to one of the refuters or defeaters (or whether his or her interlocutors are committed and entitled to one of the refuters or defeaters).

*Defeating* or *refuting* inferential relations are responsible for a claim’s being a defeater or refuter for an inferential relation but are not themselves defeaters or refuters.

---

61 If we allowed complex inferential relations, then \( s \) would be a refuter for \( p > q \) were \( s // (p > q) \) a materially good inferential relation. For simplicity’s sake I am going to largely avoid discussing the circumstances and consequences of attributing complex inferential relations.
One inferential relation $IRa$ is a *defeating* inferential relation for another inferential relation $IRb$ if and only if the premise set of $IRa$ is a defeater for $IRb$ in virtue of $IRa$. For example, if $p > q$ and $r // q$ are both materially good inferential relations then $r$ would be a defeater for $p > q$, and $r // q$ would be a defeating inferential relation for $p > q$ in virtue of its being responsible for $r$’s being a defeater for $p // q$. The same holds for *refuting* inferential relations, except that claims can also be refuted. $IRa$ is a refuting inferential relation for some claim $p$ if and only the premise set of $IRa$ is a refuter for $p$ in virtue of $IRa$. Defeating inferential relations are not themselves defeaters, nor are refuting inferential relations themselves refuters.

Having an undefeated reason provides entitlement; having an undefeated refuter removes entitlement:

If someone *has a defeasible reason* $p$ for some claim $q$ and this reason is not *defeated* then that person is *entitled* to $q$, though she could also be entitled to $q$ by some other means—independently of her being committed and entitled to $p$. If someone *has a refuter* $p$ for $q$ and this refuter is *not defeated* then that person will *not be entitled* to $q$, though she might not have been entitled to $q$ even if she were not committed to $p$.

If *one has* a defeated reason or refuter for a claim then the inferential relation between this claim and some of their commitments is defeated:

When we say that someone *has* a defeasible reason $p$ for $q$, we are not just saying that that person is committed and entitled to $p$; we are also endorsing the entitlement-preserving relation leading from $p$ to $q$ (even if the person who is committed and entitled
to \( p \) is not herself implicitly committed to that inferential relation). So when we say that someone has a defeasible reason \( p \) for \( q \) and this reason is defeated, we are not saying that the claim \( p \) or one’s commitment or entitlement to it is defeated, but rather that the materially good entitlement-preserving relation leading from \( p \) to \( q \) is defeated for that person. The same holds when we say that someone has a defeasible refuter \( p \) for \( q \) and this refuter is defeated (except that the inferential relations involved might be different).

One can have a good defeasible reason for a claim without being entitled to it:

If \( p > q \) is a materially good entitlement-preserving relation then anyone committed and entitled to \( p \) has a good defeasible reason for \( q \) regardless of what else they are committed or entitled to. If someone \( S \) is committed and entitled to \( p \), but \( p > q \) is defeated for \( S \), then \( S \) has a good defeasible reason for \( q \), even though \( S \) is not entitled to \( q \) (unless she has some other undefeated reason for \( q \)). Therefore, someone’s having a good reason for some claim is independent of whether that person is actually entitled to that claim. Whether or not one claim is a good defeasible reason for another claim is independent of what anyone is actually committed or entitled to. But, whether or not a particular person is entitled to a particular claim will depend upon the various other claims that person is committed and entitled to; what those other claims are reasons for or are reasons against; and what that particular person’s interlocutors are committed and entitled to.
Part 2: A Review of the Problem of Entitlement-Recovery

I argued in Chapter II that allowing entitlement recovery requires that defeating inferential relations be themselves defeasible. Brandom’s three types of inferential relations together do not allow entitlement recovery because:

(1) The only type that can act as defeating inferential relation are incompatibility relations and these are not themselves defeasible.

(2) The only type that are defeasible are entitlement-preserving relations and these cannot act as defeating relations because a scorekeeper can be entitled to two incompatible claims at the same time so long as she is not committed to either.

On Brandom’s account, the consequences of being committed to different types of inferential relations are to be understood in terms of score updating practices (adding and removing commitments and entitlements to various claims). This is what Brandom’s scorekeeping algorithm was intended to express with regards to assertions. The central aim of this chapter is to propose a revised scorekeeping algorithm that allows entitlement-recovery. This will primarily involve proposing a revised account of how entitlement-preserving relations can be defeated and how they can be defeating relations.

Let us, once again, use the following abbreviations on the left to represent the propositions expressed by the following sentences on the right:
\[ \text{Struck} = \text{The dry, well-made match is struck} \]

\[ \text{Lights} = \text{The match lights} \]

\[ \text{EMF} = \text{The match is in a strong electromagnetic field} \]

\[ \text{FCage} = \text{The match is in a Faraday cage} \]

\[ \text{Vacuum} = \text{The room is evacuated of oxygen} \]

In our first reconstruction of Brandom’s match-lighting scenario, we treated the negation signs in Brandom’s conditionals as indicating incompatibility relations. This gave us the following inferential relations:

\[ \text{Struck} \rightarrow \text{Lights} \]

\[ \{\text{Struck, EMF}\} // \text{Lights} \]

\[ \{\text{Struck, EMF, FCage}\} \rightarrow \text{Lights} \]

\[ \{\text{Struck, EMF, FCage, Vacuum}\} // \text{Lights} \]

The problem with this reconstruction was that \( \{\text{Struck, EMF}\} // \text{Lights} \) is indefeasible and was a defeating relation for \( \{\text{Struck, EMF, FCage}\} \rightarrow \text{Lights} \). As a result, if a scorekeeper were committed to all these inferential relations and attributed to her interlocutor commitment and entitlement to \text{Struck, EMF, and FCage} then no matter what further commitments and entitlements she attributes to this person, the scorekeeper
would be obliged to withhold attributing entitlement to \textit{Lights} (thereby preventing entitlement-recovery). Later in Chapter II, we considered reconstructing the scenario as follows

\begin{quote}
\textit{Struck} \succ \textit{Lights}
\end{quote}

\begin{quote}
\{\textit{Struck}, \textit{EMF}\} \succ \sim\textit{Lights}
\end{quote}

\begin{quote}
\{\textit{Struck}, \textit{EMF}, \textit{FCage}\} \succ \textit{Lights}
\end{quote}

\begin{quote}
\{\textit{Struck}, \textit{EMF}, \textit{FCage}, \textit{Vacuum}\} \succ \sim\textit{Lights}
\end{quote}

The problem with this reconstruction, where all the inferential relations involved are entitlement-preserving, was that we could not even get defeat. This is because scorekeepers on Brandom’s account could be entitled to \textit{q} and \sim\textit{q} at the same time, so long as they were committed to neither. A scorekeeper’s commitment to an entitlement-preserving relation can oblige her to add entitlements to whomever she is keeping score on, but they cannot oblige the scorekeeper to remove entitlements. This is why entitlement-preserving relations cannot be defeating relations.

\begin{quote}
The solution that I will propose to this problem consists of modifying entitlement-preserving relations so that they can act as defeating relations for other entitlement-preserving relations. Incompatibility relations will still be able to act as defeating relations, but they will remain indefeasible.
\end{quote}

In order to make it so that Brandom’s entitlement-preserving relations can be defeating relations for other entitlement-preserving relations they must be modified so
that (1) their premises can act as refuters for the conclusions of various other entitlement-preserving relations, (2) their premises can sometimes grant entitlement to their conclusions even in contexts where there are refuters against those conclusions, thereby allowing what I call asymmetric defeat. In the next section I will propose modifications to Brandom’s account that will allow it to meet conditions (1) and (2).

**PART 3: A Revised Account of Defeat**

Before I begin to present my revised account of defeat, I want to stress that I am not developing a system of logic for defeasible reasoning. I am concerned with *material* inferential relations, rather than the formal ones we might use to express them. My project is therefore importantly different from the one that Reiter undertakes when developing his default logic.

**Defeasible Refuters**

In order to make it so that entitlement-preserving relations can be defeating relations for other entitlement-preserving relations there must be some means by which the premises of one entitlement-preserving relation can be refuters for the conclusions of other entitlement-preserving relations. We saw that on Brandom’s account someone can be entitled to both ‘Tweety can fly’ and ‘Tweety is unable to fly’ at the same time. Therefore, claims (such as ‘Tweety is a Penguin’) that have ‘Tweety is unable to fly’ as
an entitlement-preserving consequence cannot be refuters against ‘Tweety can fly’ in virtue of its standing in that entitlement-preserving relation.

The way around this problem is simply to no longer allow scorekeepers to attribute entitlement to two or more incompatible claims to the same person at the same time (even if that person is not yet committed to any of the claims). By doing this we begin to provide a way for a claim like ‘Tweety is a penguin’ to be a refuter for other claims like ‘Tweety can fly’, without requiring that the two claims be incompatible with each other. If a scorekeeper takes ‘Tweety can fly’ to be incompatible with ‘Tweety cannot fly’, then that scorekeeper will take a reason for one of these claims to be a refuter for the other.

This understanding of refuters applies whether the reasons for the claims are defeasible or indefeasible. If $p$ is a defeasible reason for $q$ (because $q$ is an entitlement-preserving consequence of $p$) then $p$ is a defeasible refuter for any claim incompatible with $q$. If $p$ is an indefeasible reason for $q$ (because $q$ is a commitment-preserving consequence of $p$) then $p$ is an indefeasible refuter for any claim that is incompatible with $q$. The sets of claims on one side of an incompatibility relation are also indefeasible refuters against the set of claims on the other side of the relation.

Putting this altogether we can then define defeasible refuters in terms of good defeasible reasons and incompatibility.

-A claim $p$ is a defeasible refuter for some claim $q$ IF AND ONLY IF there is some claim $r$ such that $p$ is a good defeasible reason for $r$, and $r$ is incompatible
with \( q \). (Again, it does not matter whether anyone is actually committed or entitled to \( p \) or \( r \).

If, following Brandom, we treat the negation of a claim to be the intersection of the sets of claims that are incompatible with every claim incompatible with \( p \) then we can use entitlement-preserving relations with negated conclusions to express defeasible refuters. For example, \( p > \sim q \) can be interpreted as saying that \( p \) is a defeasible refuter for \( q \).

While we will have to wait until later sections to fully spell out the conditions in which a good defeasible reason is defeated, we can presently identify when a defeasible refuter is defeated in terms of when good defeasible reasons are defeated.

\[
\text{IF} \quad p \quad \text{is a refuter for} \quad q \quad \text{in virtue of} \\
\quad p’\text{’s being a good defeasible reason for} \quad r \\
\quad r’\text{’s being incompatible with} \quad q \\
\text{THEN} \\
\text{IF} \quad \text{a reasoner’s defeasible reason} \quad p \quad \text{for} \quad r \quad \text{is defeated} \\
\text{THEN} \quad \text{The reasoner’s defeasible refuter} \quad p \quad \text{for} \quad q \quad \text{is defeated}
\]

**Having a Reason vs Being Entitled**

Before carrying on, I will address a confusion which likely motivates Brandom’s claiming that one can be entitled to two or more incompatible claims at the same time. The confusion rests on conflating someone’s having a good reason to undertake some
commitment \( p \), and someone’s \textit{being entitled} to undertake some commitment \( p \). For if we conflate these then we would be obliged to endorse the inference from ‘reasons game players can have \textit{good reasons} for two incompatible claims’ to ‘reasons game players can be \textit{entitled} to two incompatible claims’.

It is true that if someone is entitled to claim \( p \) then that person will probably have a good reason for \( p \) (supposing we ignore default entitlement and non-inferential observation reports). However, if we are to take defeasible inference seriously then we cannot agree that the inverse holds. If the inferential relation leading from ‘Tweety is a bird’ to ‘Tweety can fly’ is materially good then it can remain a good inference even when commitment and entitlement to ‘Tweety is a bird’ does not succeed in providing entitlement to ‘Tweety can fly’. This shows that someone can have a good reason for ‘Tweety can fly’ even if she is not in fact entitled to that claim. As a result, there is an important distinction between the claim that I am endorsing—that reasons game players are able to have \textit{good reasons} for two or more incompatible claims—and the claim that I am denying—that reasons game players can be \textit{entitled} to two or more incompatible claims.

As discussed, when we say that a scorekeeper is implicitly committed to a defeasible inferential relation leading from \( p \) to \( q \) we are not saying that the scorekeeper’s \textit{being committed} to \( p \)’s being a good reason for \( q \) is defeasible in the sense of being revisable. We are instead saying that the type of inferential relation that the scorekeeper takes to hold between \( p \) and \( q \) is such that she continues to take \( p \) to be a good reason for \( q \) even in those contexts where she will not draw conclusion \( q \) from premise \( p \). One of the challenges in thinking about defeasible reasoning is getting oneself to stop thinking of
‘good reasons’ in purely deductive terms where commitment and entitlement to the premises guarantees entitlement to the conclusion. To avoid confusion it is often helpful to say premise p is a good defeasible reason for conclusion q, where ‘defeasible’ indicates what kind of reason it is rather than indicating that p’s being a good reason for q is defeasible. If we restricted our notion of ‘good reason’ to the sort where the premises guarantee entitlement to the conclusion then we would no longer be discussing defeasible reasoning.

Asymmetric Defeat

Above I showed how we can modify Brandom’s conception of entitlement and entitlement-preserving relations so that the premises of entitlement preserving relations can act as refuters against the conclusions of various other entitlement-preserving relations. I will now show how we can further modify Brandom’s entitlement-preserving relations so that their premises can sometimes grant entitlement to their conclusions in contexts where there are refuters against these conclusions. This will allow what I call asymmetric defeat.

Keeping in mind the distinction between being entitled to some claim and having a good reason for that claim, we must now answer the question, ‘If a reasoner has good defeasible reasons for two incompatible claims, which (if either) of those claims is the reasoner entitled to undertake commitment to?’ I claimed towards the end of Part I that one is entitled to any claim that one has a good undefeated reason for. This lets us
usefully reformulate the question as follows, ‘If a reasoner has good defeasible reasons for two incompatible claims, which (if either) of those reasons is undefeated’.

One candidate answer is that any good defeasible reason for a claim $q$ is defeated for anyone who has any refuter for $q$—when one is committed to something incompatible with $q$ or has a good reason for some claim that is incompatible with $q$. So if $p$ is a good defeasible reason for $q$, and $r$ is a good defeasible reason for $s$, but $q$ is incompatible with $s$, then anyone committed to both $p$ and $r$ cannot be entitled to either $q$ or $s$. This sort of symmetric (or mutual) defeat is appropriate in situations where a reasoner should remain agnostic about $q$ and $s$ because of her having competing reasons/refuters; but there are also many situations (including those that involve entitlement-recovery) where defeat should be asymmetrical, whereby someone can be entitled to $q$ despite her having a refuter for it. For example, ‘Tweety is dead’ is a much stronger reason for ‘Tweety is unable to fly’ than ‘Tweety is a bird’ is a reason for ‘Tweety can fly’, but ‘Tweety is a bird’ is nonetheless a refuter for ‘Tweety is unable to fly’. Therefore commitment and entitlement to ‘Tweety is dead’ should defeat the inference from ‘Tweety is a bird’ to Tweety can fly’, but commitment and entitlement to ‘Tweety is a bird’ should not defeat the inference from ‘Tweety is dead’ to ‘Tweety cannot fly’. This shows that it is not the case that any good defeasible reason for a claim $q$ is defeated any time someone is committed to a refuter for $q$. Therefore, an inferential relation’s being defeated for someone requires that some further condition be satisfied.
I propose that the additional condition required for defeat reflect some version of what is commonly called ‘the principle of specificity’ in the AI literature.\(^{62}\) This principle, which can take various different forms, roughly states that if one accepts the premises of two or more competing defeasible inference relations (often like Reiter-style defaults) then one should give preference or priority to the inferential relation that incorporates more specific evidence into its premises. So if one accepts

\begin{itemize}
  \item \textit{Tweety is a penguin}
  \item \textit{Tweety is a bird} \implies \textit{Tweety can fly}
  \item \textit{Tweety is a penguin} \implies \textit{Tweety cannot fly}
  \item \textit{Tweety is a penguin} \implies \textit{Tweety is a bird}
\end{itemize}

then it prima facie looks as though one could conclude either ‘Tweety can fly’ or ‘Tweety cannot fly’. But because ‘Tweety is a penguin’ is more specific than ‘Tweety is a bird’, one should prefer the inference from ‘Tweety is a penguin’ over the inference leading from ‘Tweety is a bird’ and therefore conclude that ‘Tweety cannot fly’, but not ‘Tweety can fly’.\(^{63}\) In the AI literature there are many different accounts as to how we are to assess which inferential relations incorporate more specific evidence in their premises. But given that they are dealing specifically with a sort of formal validity rather than material goodness, they are in many ways dealing with a different sort of problem than we are dealing with here.

\(^{62}\) See, for example, Poole (1985)
\(^{63}\) As it stands, the premise ‘Tweety is dead’ (from the previous example) is not more specific than the premise ‘Tweety is a bird’ in the same sense that ‘Tweety is a penguin’ is more specific than ‘Tweety is a Bird’. But, as we shall see momentarily, this is not a problem.
The version of the principle of specificity that I intend to use as the additional condition for defeat is remarkably simple but effective.

Some premise set A is more specific than another premise set B IFF B is a proper subset of A. IF someone accepts the premises of two competing defeasible inferential relations IRa and IRb AND the set of premises belonging to IRa is more specific than the premise set belong to IRb THEN IRa should be given preference over IRb, such that IRa is a defeating relation for IRb, but IRb is not a defeating relation for IRa.

We now have an answer to the question, ‘If a reasoner has good defeasible reasons for two incompatible claims, which (if either) of those reasons is undefeated?’ Let IRx be any inferential relation belonging to some set of competing inferential relations. IRx is undefeated IF the premise sets belonging to each of the inferential relations competing with IRx are proper subsets of IRx’s premise set.

I will now present two different examples. The first is an example of symmetric defeat, while the second is an example of asymmetric defeat. Suppose that we have the following inferential relations:

IR1 {Today is the last Saturday before Christmas} > The mall will be busy today

IR2 {There is a snow storm today} > The mall will be quiet today

IR3 The mall will be quiet today // The mall will be busy today
Because the conclusion of IR1 and IR2 are incompatible with one another, the premise set of IR1 is a refuter for the conclusion of IR2 and the premise set of IR2 is a refuter for the conclusion of IR1. Therefore, ‘Today is the last Saturday before Christmas’ and ‘There is a snow storm today’ are competing reasons with regards to both ‘The mall will be quiet today’ and ‘The mall will be busy today’. Because the premise set of IR1 is not a proper subset of the premise set of IR2 and the premise set of IR2 is not a proper subset of the premise set of IR1, if someone S were committed and entitled to both premise sets then both IR1 and IR2 would be defeated for S. This would be an instance of symmetric defeat, whereby S would therefore not be entitled to either ‘The mall will be busy today’ or ‘The mall will not be busy today’ unless she had some other good but undefeated reason for one of these claims.

For the second example let us now suppose that ‘Today is the last Saturday before Christmas’ is a much stronger reason for the conclusion ‘The mall will be busy today’ than ‘There is a snow storm today’ is a refuter for this conclusion. On this supposition, S’s commitment and entitlement to ‘Today is the last Saturday before Christmas’ would entitle S to ‘The mall will be busy today’ even if S were also committed and entitled to ‘There is a snow storm today’. The entitlement-preserving relation that captures this and that must be added to the set of inferential relations above is

\[ IR4 \{ \text{Today is the last Saturday before Christmas, There is a snow storm today} \} \succ \text{The mall will be busy today} \]
Because the conclusion of \textit{IR4} is incompatible with the conclusion of \textit{IR2}, the premise sets of \textit{IR4} and \textit{IR2} are competing reasons for ‘The mall will be busy today’ and ‘The mall will be quiet today’. However, even though their premise sets provide competing reasons, it is not the case the \textit{IR2} and \textit{IR4} are mutually defeating inferential relations. Because the premises set of \textit{IR2} is a proper subset of the premise set of \textit{IR4}, \textit{IR4} is a defeating inferential relation for \textit{IR2}, but \textit{IR2} is not a defeating inferential relation for \textit{IR4}. Therefore, if S were committed and entitled to the premises in \textit{IR4} and \textit{IR2} then this would be an instance of asymmetric defeat, whereby \textit{IR2} is defeated for S but \textit{IR4} is not (assuming that S does not have some other undefeated refuter for ‘The mall will be busy’). Because S has an undefeated reason for the claim ‘The mall will be busy today’ but no undefeated reason for the claim ‘The mall will be quiet today’, it follows that S is entitled to the former claim, but not the latter.\footnote{Despite lacking entitlement to ‘The mall will be quiet today’, S still has a good defeasible reason for it.}

In this example of asymmetric defeat we see that \textit{IR2} remains a defeating inferential relation for \textit{IR1} even though \textit{IR2} is itself defeated. Furthermore, \textit{IR1} is still defeated for S even though it turns out that S is entitled to \textit{IR1}’s conclusion (in virtue of her being committed and entitled to the premises of \textit{IR4}). An entitlement-preserving relation’s being defeated for some person S is independent of whether or not S is entitled to the conclusion of that inferential relation, much in the same way that a deductive argument’s validity is independent of the truth of its conclusion.

When we added \textit{IR4} to the example of symmetric defeat, there may be a temptation to have \textit{IR4} replace \textit{IR1} because \textit{IR1} does not seem to do any work in the
example where we consider asymmetric defeat. The reason why we do not replace \textit{IRI} is because its inclusion tells us that the claim ‘Today is the last Saturday before Christmas’ is by itself a good defeasible reason for ‘The mall will be busy today’. So if one were committed and entitled to ‘Today is the last Saturday before Christmas’, but not committed to ‘There is a snow storm today’ then one would be consequentially entitled to ‘The mall will be busy today’ (unless one has some other undefeated refuter for that conclusion).

\textbf{Interpersonal Defeat}

On Brandom’s original account of the reasons game, the players can lose entitlement to claims as a result of challenges that are made by their interlocutors. As I explained in Chapter I and later in Chapter IV, reasons game players have a responsibility to justify their assertional commitments when appropriately challenged. Once a player’s entitlement to undertake commitment to claim \( p \) is appropriately challenged she loses entitlement to \( p \) unless and until she can demonstrate entitlement to it (either by asserting premises from which \( p \) follows or by deferring justificatory responsibility to someone else). What is important about this for present purposes is that Brandom goes on to suggest that we can treat challenges to entitlements as themselves assertions.

Indeed, the simplest way to implement such a feature of the model of asserting [where challenges may themselves need justification] is to require that the performances that have the significances of challenging entitlements to assertional commitments be themselves assertions. One can
then challenge an assertion *only* by making an assertion incompatible with it. Recall that two claims are incompatible just in case commitment to one precludes entitlement to the other. Then challenges have no privileged status: their entitlement is on the table along with that of what they challenge. (MIE 178, Italics added)

Let us consider an example where ‘Tweety can fly’ is an entitlement-preserving consequence of ‘Tweety is a bird’ and (temporarily sticking to Brandom’s original account) ‘Tweety is a penguin’ is incompatible with ‘Tweety can fly’. Let us now suppose that reasoner S1 is not committed to ‘Tweety is a penguin’, but is committed and entitled to ‘Tweety is a bird’ and uses this to justify her entitlement to ‘Tweety can fly’. Reasoner S2 then challenges S1’s entitlement to ‘Tweety can fly’ by asserting ‘Tweety is a penguin’. Here is the crucial bit: if reasoner S2 is entitled to ‘Tweety is a penguin’ then (on Brandom’s account) S1’s entitlement to ‘Tweety can fly’ will be successfully refuted and the entitlement preserving relation from ‘Tweety is a bird’ to ‘Tweety can fly’ will be defeated for S1 *even though S1 is not herself committed to the refuting/defeating claim*.

What this shows is that if refutation can occur interpersonally, which it must if reasoners can challenge the entitlements of others, then defeat must also occur interpersonally (as seen in the example). This means that whether or not a defeasible inferential relation is defeated for some reasoner S depends not only on S’s score, but on S’s interlocutors’ scores as well.
The other important lesson to be drawn from Brandom’s discussion of challenges to entitlement is that a reasoner’s merely being committed to a refuter for \( p \) is not sufficient to successfully refute an interlocutor’s entitlement to \( p \). The challenger must be both committed and entitled to the refuter if she is to successfully refute an interlocutor’s entitlement. This is in contrast to the situation where the reasoner refutes one of her own entitlements, whereby she need not be entitled to the refuter. On Brandom’s original and my revised account, it is possible (though not permissible) for reasoners to be committed to two incompatible claims. When this happens reasoners cannot be entitled to either one of the claims unless they disavow commitment to the other claim. Because they are not entitled to either of the incompatible claims that they are committed to, they cannot use either of these claims to successfully challenge/refute the entitlements of their interlocutors.

Given that the only way that one claim can be a refuter for another claim on Brandom’s account is for the two to be incompatible with one another, it is obvious why Brandom says that one can only challenge an assertion by making another assertion that is incompatible with the first. However, on the revised account of defeasible reasoning being developed here, there is another way that a claim \( p \) can be a refuter for another claim \( q \) (and hence another way that asserting \( p \) can challenge someone’s entitlement to \( q \)) : \( p \) is a defeasible refuter for \( q \) if \( p \) has an entitlement-preserving consequence that is incompatible with \( q \). The defeasible refuters on my revised account function almost identically to Brandom’s indefeasible refuters (which continue to be part of my revised account) except that they have the added defeasibility conditions.
Regardless of whether $p$ is a defeasible or indefeasible refuter for $q$:

- a reasoner need not be entitled to claim $p$ in order for her entitlement to $q$ to be refuted by her own commitment to $p$.

- a reasoner must be entitled to $p$ if her commitment to $p$ is to refute her interlocutor’s entitlement to $q$.

When Defeasible Reasons Compete with Indefeasible Reasons

I have been focusing upon how we are to answer the question “If a reasoner has good defeasible reasons for two incompatible claims, which (if either) of those claims is the reasoner entitled to undertake commitment to?”. I have not addressed the question ‘If a reasoner has a good defeasible reason for one claim and a good indefeasible reason for another claim and the two claims that she has reasons for are incompatible, which if either of those claims is the reasoner entitled to?’ I will only briefly address this latter question because it is so much easier to answer. Any time the conclusion of a defeasible inferential relation is incompatible with the conclusion of an indefeasible inferential relation (or one side of an incompatibility relation) then the indefeasible inferential relation is a defeating relation for the defeasible relation. As should be obvious, defeasible inferential relations are never defeating inferential relations for indefeasible inferential relations.
Circumstances of Defeat:

Thus far I briefly reviewed the problem of entitlement-recovery and then suggested that one way to overcome it is to modify entitlement-preserving relations so that they can be defeating relations for other entitlement-preserving relations. I then began working out the details of this suggestion by proposing a modification that makes it so that the premises of one entitlement-preserving relation can be defeasible refuters for the conclusions of other entitlement-preserving relations. From there I proposed some conditions for defeat that allow asymmetric defeat, thereby allowing scorekeepers to be entitled to claims even if they have refuters against those claims.

Building on the preceding discussion, I will now present the scorekeeping circumstances in which a scorekeeper is obliged to take an inferential relation to be defeated, while avoiding using any expressions like ‘reasons for’, or ‘refuters against’ that are not part of the basic scorekeeping account (or the three types of inferential relations). This task must not be confused with the one of identifying the circumstances for appropriately attributing inferential commitments, which will be addressed in the next chapter. In order to maintain the defeat/refute distinction, the conditions in which it is appropriate to attribute a commitment to a defeasible inferential relation must be kept independent of the conditions in which someone with that inferential commitment would be obliged to take it to be defeated.

I explained earlier that when a scorekeeper updates the scores she attributes to her interlocutors, she uses her own inferential commitments rather than those of her interlocutors. As a result, each scorekeeper takes herself and each of her interlocutors to
be bound by the same set of inferential relations, even though different scorekeepers will not entirely agree on which set of inferential relations apply to everyone. Defeat, however, does not work this way. A scorekeeper $S_1$ may take some defeasible inferential relation to be defeated for her interlocutor $S_2$ but not defeated for her interlocutor $S_3$ in virtue of $S_2$, but not $S_3$, being committed to some claim that defeats $p > q$ (or having some interlocutors committed and entitled to some claim that defeats $p > q$).

A scorekeeper $S_1$ is obliged to take entitlement-preserving relation $A > p$ to be **defeated** for $S_2$ IF $S_1$ is implicitly committed to $A > p$ AND

There exist some set of claims $X$ and some claim $y$ such that
- $S_1$ is implicitly committed to $X > y$  
  (S1 takes $X$ to be a defeasible refuter for $p$)
- $S_1$ is implicitly committed to $y // p$  
  (Premises A not more specific than X)
- $X$ is not a proper subset of $A$  
  (S1 takes $S_2$ to have refuter $X$ for $p$)
- S1 attributes commitment $X$ to $S_2$  
  (OR S1 attributes commitment and entitlement $X$ to $S_2$’s interlocutors)

OR

There exists some set of claims $X$ such that
- $S_1$ is implicitly committed to $X // p$  
  (S1 takes $X$ to be a refuter for $p$)
- S1 attributes commitment $X$ to $S_2$  
  (OR S attributes commitment and entitlement $X$ to $S_2$’s interlocutors)

OR

There exists some set of claims $X$ and some claim $y$ such that
- $S_1$ is implicitly committed to $X \rightarrow y$  
  (S1 takes $X$ to be an indefeasible refuter against $p$)
- $S_1$ is implicitly committed to $y // p$  
  (S1 takes $S_2$ to have refuter $X$ for $p$)
- S1 attributes entitlement $X$ to $S_2$  
  (OR S1 attributes commitment and entitlement $X$ to $S_2$’s interlocutors)
The first disjunct above specifies the conditions in which an entitlement-preserving relation is the defeating relation for $A > p$. Notice that S1 need not attribute entitlement $X$ to S2 in order for S2’s commitment to $X$ to be a refuter (or a defeater). The second and third disjuncts specify when incompatibility relations and commitment-preserving relations are the defeating relations.

Consequences of Defeat

The consequence of scorekeeper S1’s taking $A > p$ to be defeated for S2 is that S1 is not obliged to attribute entitlement $p$ to S2 in response to attributing to S2 commitment and entitlement to $A$. S1 might still, however, attribute some other commitments and entitlements to S2 that oblige S1 to attribute entitlement $p$ to S2. Furthermore, as discussed above, a defeated entitlement-preserving relation can still be a defeating inferential relation and can therefore still make an important difference to how a scorekeeper should update score. Further details about the consequences of defeat will be spelled out in the revised scorekeeping algorithm below.

**PART 4: Revising the Scorekeeping Algorithm**

I will now show how we can incorporate the revised account of defeat into a revised scorekeeping algorithm that allows entitlement-recovery. When the defeating inferential relation for an entitlement-preserving relation $A > p$ is itself indefeasible,
because it is a commitment-preserving or incompatibility relation, then entitlement to \( p \) cannot be recovered. But if the defeating inferential relation is defeasible, because it is an entitlement-preserving relation, then entitlement-recovery is possible. However, in order to show this I must explain how scorekeepers committed to different types of inferential relations are obliged to update score. To do this I will now propose a revised version of the scorekeeping algorithm. The original scorekeeping algorithm (found in MIE 190-191) was as follows,

**Brandom’s Original Scorekeeping Algorithm**

How scorekeeper S1 is obliged to update score in response to S2’s asserting \( p \):

**Step 1: Add commitment \( p \)**

To begin with, [S1] must add \( p \) to the list of commitments attributed to [S2] (supposing the more interesting case in which [S1] does not already attribute to [S2] a commitment with that content).(MIE 190)

**Step 2: Add commitments to commitment-preserving consequences of \( p \)**

[S2] should then add also commitment to any claims \( q \) that are committive-inferential consequences of \( p \), in the context of the other claims attributed to [S2]. These will vary depending on the auxiliary
hypotheses available, according to what other commitments [S1] already attributes to [S2]. (MIE 190-191)

**Step 3: Remove entitlements that are incompatible with p**

Next the incompatibility relations that p (and so its commitment-inferential consequences) stand in must be consulted to determine which, if any, of the entitlements [S1] previously attributed to [S2] are precluded by the newly attributed commitment. Assertions add new commitments, but they can not only add but also subtract entitlements. (MIE 191)

**Step 4: Add entitlements to the conclusions of any undefeated commitment-preserving relations whose premises S2 is already entitled**

Then in light of the incompatibility relations associated with all of the commitments attributed to [S2], [S1] can attribute entitlements to any claims that are committive-inferential consequences of commitments to which [S2] is already taken to be entitled, closing the score under commitment-preserving inferences, where the resulting attributions of entitlement are not defeated by incompatibilities. [Brandom’s endnote: “If one is entitled to p and p commitment-entails q, one is entitled to q—any entitlement-defeating incompatibilities to q equally defeat entitlement to p. (MIE 674 note 43).] (MIE 191)
Step 5: Add entitlements to the conclusions of any undefeated entitlement-preserving relations whose premises S2 is already committed and entitled

Next constrained by the entitlement-precluding incompatibility relations associated with all of the other commitments attributed to [S2], [S1] can attribute entitlements to any claims that are permissive-inferential consequences of commitments to which [S2] is already taken to be entitled, closing the attributed score under entitlement-preserving inferences not defeated by those incompatibilities.(MIE 191)

Step 6: IF applicable, add entitlement p to S2’s interlocutors

Then [S1] needs to assess [S2]’s entitlement to the claim that p, by looking at good inferences having it as a conclusion and premises to which [S2] is committed and entitled… If [S1] takes [S2] to be entitled to p by [commitment or entitlement preserving relations, by testimony, by default, or by reliability inferences] then [S1] will take [S2] to successfully entitle others (including [S1]) to that claim (in the absence of incompatible defeasors).(MIE 191)
Required Revisions for the Original Scorekeeping Algorithm

There is no need to revise the first three steps in Brandom’s scorekeeping algorithm. The way that commitment-preserving and incompatibility relations behave in my revised scorekeeping algorithm is nearly identical to the way that they behave in Brandom’s original.

The most significant change that I make to the algorithm is the addition of another step just after Brandom’s Step 3. This new step will require that S1 remove any entitlements that S2 has a refuter for, regardless of whether the associated refuting relations are defeated or not. (Any entitlements that are removed because of S2’s being committed to the premises of a defeated refuting relation will be re-added in a later scorekeeping step). It is necessary to add this step (the one to be added just after Brandom’s Step 3) because Brandom’s original scorekeeping algorithm only has a step for removing entitlements that S2 has an indefeasible refuter for. But, in order to allow entitlement recovery S1 must also be obliged remove entitlements that S2 has a defeasible refuter for. As explained above, p can be a defeasible refuter for q in virtue of its having some claim incompatible with q as an entitlement-preserving consequent.

Brandom’s Step 4 remains largely unchanged in the revised algorithm, but its position in the revised algorithm is moved so that it is comes after the step corresponding to Brandom’s Step 5. This is necessary because the newly added step (just discussed) will likely have scorekeeper S1 temporarily remove some of S2’s entitlements that will need to be re-added before S1 goes on to add those entitlements that are commitment-preserving consequences of claims that S2 is already entitled to. For example, suppose
that S1 takes S2 to be entitled to \( q \) prior to S2’s asserting \( p \). Also suppose S1 takes \( p \) to be a defeasible refuter for \( q \), and takes the refuting inferential relation responsible for this to be defeated for S2. When S1 updates S2’s score according to the newly added step (the one to be inserted just after Brandom’s Step 3), S1 will remove entitlement \( q \) even though S2’s refuter for \( q \) is defeated for S2. At some later step in the algorithm entitlement \( q \) will have to be re-added to S2’s score, but if this happens only after S1 adds entitlements that are commitment-preserving consequences of S1’s prior entitlements, then the algorithm will not have S1 add entitlements that are commitment-preserving consequences of entitlement \( q \).

Brandom’s Step 5 also needs to be revised. As it stands, it only says that scorekeepers should refrain from adding entitlement-preserving consequences that are not defeated by incompatibility relations. It must therefore be modified so that it will involve defeating relations that are not incompatibility relations. Thankfully most of the work has already been achieved in the revised account of defeat presented above. Once this is done we need to remove any reference to entitlements being defeated. On the revised account, it makes no sense to say that commitments, entitlements, or claims can be defeated; only inferential relations can be defeated. The modified version of Step 5 ends up saying “S1 must add entitlements to any claims that are conclusions of undefeated entitlement-preserving relations whose premises S2 is both committed and entitled to and which \( p \) (or some other newly added entitlement) is a member of. S1 repeats this as necessary for each entitlement added.” I add the ‘repeat as necessary’ instruction because a newly added claim \( p \)’s entitlement-preserving consequences will themselves have entitlement-preserving consequences that \textit{may} have to be added (such as in cases where S2 was
committed but not entitled to p’s entitlement-preserving consequences prior to asserting p). It is for simplicity’s sake that I only refer to the revised account of defeat (presented above in part 3) rather than fully spell it out within the scorekeeping algorithm.

The last change that I make to Brandom’s original scorekeeping algorithm consists of slightly modifying his step 6 to allow for defeasible refuters against p. Again, Brandom only considers refuters that are incompatible with p (and which are therefore indefeasible) whereas in the revised account we must consider refuters against p that are defeasible.

**The Revised Scorekeeping Algorithm:**

Below is the revised scorekeeping algorithm. As it was with Brandom’s original scorekeeping algorithm, all inferential relations appealed to in the revised scorekeeping algorithm are those that the scorekeeper S1 is implicitly committed to rather than those that S2 (the person she is keeping score on) is committed to. Once again S1 is updating S2’s score in response to S2’s asserting p (and again we are assuming that S2 was not already committed to p).

**Revised Step 1: Add commitment p**

S1 must add p to S2’s commitments

**Revised Step 2: Add commitments to claims that p is an indefeasible reason for**
S1 must add commitments to claims that are the conclusions of any commitment-preserving inferential relations whose premises S2 is committed to and that \( p \) is a member of.

Revised Step 3: Remove entitlements that \( p \) is an indefeasible refuter for

S1 must remove entitlement to any claims that stand in incompatibility relations with any subset of S2’s commitments that \( p \) is a member of.

Revised Step 4: Remove entitlements that \( p \) is a defeasible refuter for

S1 must remove entitlement to any claims that are incompatible with the conclusions of any entitlement-preserving relations whose premises S2 is committed to and that \( p \) is a member of.

Revised Step 5: Add entitlements to any claims that S2 has an undefeated defeasible reason for

S1 must add entitlements to any claims that are conclusions of undefeated entitlement-preserving relations whose premises S2 is both committed and entitled to and which \( p \) (or some other newly added entitlement) is a member of. S1 repeats this as necessary for each entitlement added.

Revised Step 6: Add entitlements to any claims that S2 has an indefeasible reason for

S1 must add entitlements to any claims that are conclusions of commitment-preserving relations whose premises S2 is entitled to.
Revised Step 7: If S2 is entitled to $p$ then add entitlement $p$ to each of S2’s interlocutors that does not have an undefeated refuter for $p$

IF S1 takes S2 to be entitled to $p$ then S1 must add $p$ to the entitlements of each of S2’s interlocutors that is neither committed to any set of claims incompatible with $p$ nor committed to any set of claims that is a premise set of an entitlement-preserving relation to whose conclusion is incompatible with $p$ and which is undefeated for that interlocutor.

**PART 5: Putting to Work the Revised Account of the Consequences of Inferential Commitment**

I will now show what we can do with the revised scorekeeping algorithm and the revised account of defeat that it incorporates. First, and most importantly, I will show that it allows entitlement-recovery. After that, I will go on to show how we can use the revised account of the consequences of inferential commitment to understand the defeasibility involved in default entitlement. Lastly, I will very briefly explain how we can use the revised account to understand refuters against inferential relations and how the set of inferential commitments that make a claim a refuter for an inferential relation are distinct from those that make it a defeater for that inferential relation.
The Revised Scorekeeping Algorithm Allows Entitlement-Recovery

Now comes the moment of truth. I will show how the revised scorekeeping algorithm, incorporating the revised account of defeat, obliges the scorekeeper in the match-lighting scenario to update score correctly. I will use the same abbreviations that I previously used but will also introduce one for the claim ‘The match remains unlit’. This is just so that we do not have to assume formal negation. Under the ‘Explanation’ column I will identify which step in the revised scorekeeping algorithm and which inferential commitment is responsible for the indicated score update.

*Struck* = The dry, well-made match is struck

*Lights* = The match lights

*Unlit* = The match remains unlit

*EMF* = The match is in a strong electromagnetic field

*FCage* = The match is in a Faraday cage

*Vacuum* = The room is evacuated of oxygen
TABLE 4-1
Updating Score in the Match-lighting Example
According to the Revised Scorekeeping Algorithm

S1’s Inferential Commitments:

IR-A) \{Struck\} \rightarrow \text{Lights}
IR-B) \{Struck, EMF\} \rightarrow \text{Unlit}
IR-C) \{Struck, EMF, FCage\} \rightarrow \text{Lights}
IR-D) \{Struck, EMF, FCage, Vacuum\} \rightarrow \text{Unlit}
IR-E) \text{Lights} // \text{Unlit}

<table>
<thead>
<tr>
<th>S2’s Score (According to S1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Claims</strong></td>
</tr>
<tr>
<td>STAGE 1</td>
</tr>
<tr>
<td><strong>Initial Score:</strong></td>
</tr>
<tr>
<td>Struck</td>
</tr>
<tr>
<td>Lights</td>
</tr>
<tr>
<td><strong>Move Made:</strong></td>
</tr>
<tr>
<td>EMF</td>
</tr>
<tr>
<td><strong>Score Updated:</strong></td>
</tr>
<tr>
<td>Lights</td>
</tr>
<tr>
<td>Unlit</td>
</tr>
<tr>
<td>STAGE 2</td>
</tr>
<tr>
<td><strong>Initial Score:</strong></td>
</tr>
<tr>
<td>Struck</td>
</tr>
<tr>
<td>EMF</td>
</tr>
<tr>
<td>Unlit</td>
</tr>
<tr>
<td><strong>Move Made:</strong></td>
</tr>
<tr>
<td>FCage</td>
</tr>
<tr>
<td><strong>Score Updated:</strong></td>
</tr>
<tr>
<td>Unlit</td>
</tr>
<tr>
<td>Lights</td>
</tr>
<tr>
<td>STAGE 3</td>
</tr>
<tr>
<td><strong>Initial Score:</strong></td>
</tr>
<tr>
<td>Struck</td>
</tr>
<tr>
<td>Lights</td>
</tr>
<tr>
<td>EMF</td>
</tr>
<tr>
<td>FCage</td>
</tr>
<tr>
<td><strong>Move Made:</strong></td>
</tr>
<tr>
<td>Vacuum</td>
</tr>
<tr>
<td><strong>Score Updated:</strong></td>
</tr>
<tr>
<td>Lights</td>
</tr>
<tr>
<td>Unlit</td>
</tr>
</tbody>
</table>
What this score updating table shows is that if a scorekeeper were implicitly committed to the indicated inferential relations and were keeping score on someone whom she took to be have the indicated commitments and entitlements then she would be obliged to correctly update score in the match-lighting scenario. In particular, the score updating table shows that the revised scorekeeping algorithm allows nonmonotonic reasoning (where a claim \( \text{Lights} \) can be inferred from one set of premises \( \{ \text{Struck} \} \), but not a superset of it \( \{ \text{Struck, EMF} \} \) and entitlement recovery (where there is a further superset of premises \( \{ \text{Struck, EMF, FCage} \} \) from which \( \text{Lights} \) can once again be inferred). At each stage in reasoning the scorekeeper would alternate between attributing \( \text{Lights} \) and \( \text{Unlit} \) to her interlocutor. At no point does the scorekeeper attribute to her interlocutor entitlement to both \( \text{Lights} \) and \( \text{Unlit} \).

**Understanding Default Entitlement in Terms of the Revised Entitlement-Preserving Relations**

According to Brandom’s model we are entitled to undertake certain commitments in the reasons game by default, and if an interlocutor wants to challenge something that we have default entitlement to then the burden is upon the interlocutor to justify their
challenge. For example, if an interlocutor wants to challenge one’s default entitlement to claims like ‘Some sheep are black’ then the interlocutor cannot just ask ‘How do you know that?’ She must provide a reason (assert something) that undermines entitlement to the claim she wants to challenge, for example ‘what appears to be black sheep wool is actually dark brown when looked at closely’. This prevents us from having infinitely long chains of justification without holding the foundationalist view that entitlement to some claims cannot be challenged.

I will now show how we can understand default entitlements and challenges to them in terms of my revised entitlement-preserving relations and the corresponding account of how they are defeated. Default entitlement to \(p\) can be expressed as ‘\(\{\} \rightarrow p\)’, whereby the set brackets really do just specify an empty set of premises. If \(\{\} \rightarrow p\) is a good entitlement-preserving relation then everyone who is not committed to some refuter for \(p\) (or for \(\{\} \rightarrow p\)) will be entitled to \(p\).

One can challenge one’s interlocutors default entitlement to \(p\) by asserting (and being entitled to assert) any claim that defeats \(\{\} \rightarrow p\) according to the same account of defeat that is used for other entitlement-preserving relations. On Brandom’s original account this could only be done by asserting some claim that is incompatible with \(p\). My revised account opens new options by allowing one to challenge one’s interlocutor’s default entitlement to \(p\) by asserting a set of premises that has some claim incompatible with \(p\) as a conclusion.

For example, if the following inferential relations hold \(\{\} \rightarrow p, q \rightarrow r, r \rightarrow p\) then asserting (and being entitled to assert) \(q\) or \(r\) would defeat one’s interlocutor’s
default entitlement to \( p \). If one asserts \( r \) then one asserts an indefeasible refuter for \( p \) thereby defeating \{ \} \succ p. If instead one asserts \( q \) then one does not assert something incompatible with \( p \), but one still asserts a defeater for \{ \} \succ p by asserting a defeasible refuter for \( p \) (in virtue of \( q \) having entitlement-preserving consequence \( r \), which is incompatible with \( p \)). \( q \succ r \) is a defeating relation for \{ \} \succ p but \{ \} \succ p is not a defeating relation for \( q \succ r \) because their conclusions are incompatible and the premises of \{ \} \succ p is a proper subset of the premises of \( q \succ r \).

**Conclusion:**

In this chapter, I clarified how I was going to use some of the vocabulary that is commonly involved in discussions about defeasible reasoning. I then reviewed the problem of entitlement-recovery and proposed a revised account of defeat that plays a crucial role in overcoming this problem. The revised account of defeat was then used in the revised scorekeeping algorithm, which constitutes how a scorekeeper’s inferential commitments obliged her to update score when someone makes an assertion in the reasons game. Lastly, I showed that the revised scorekeeping algorithm allows entitlement-recovery, showing how a scorekeeper following it would update score correctly in Brandom’s match-lighting scenario.

However, an adequate account of defeasible reasoning must do more than allow entitlement-recovery. As I argued in Chapter I, and more extensively in Chapter III, an adequate account of defeasible reasoning must also distinguish an inferential relation’s being defeated from its being refuted. Therefore, an account of defeasible reasoning must
also come with an account of the circumstances in which a scorekeeper accepts different types of inferential relation, whereby the circumstances in which a scorekeeper takes a defeasible inferential relation to be defeated for herself cannot be a subset of those circumstances where the scorekeeper takes the inferential relation to be refuted. This problem will be dealt with in the next chapter.
CHAPTER VI:

Solving the Problem of Defeaters that are not Refuters:

The Revised Account of the Appropriate Circumstances for Attributing the

Three Different Types of Material Inferential Relations

PART 1: The Challenge at Hand

In the last chapter I presented my revised account of the consequences of one’s being committed to each of the three different kinds of inferential relations. This let us solve the problem of entitlement recovery. In this chapter I will present my revised account of the appropriate circumstances for attributing inferential commitments in an attempt to overcome the problem of defeaters that are not refuters.

Review of the Problem

As I explained in Chapter III, scorekeepers on Brandom’s account cannot be interpreted as committed to inferential relations simply in virtue of their being disposed to assert sentences expressing inferential relations. Brandom has us assume that the scorekeepers that we are interpreting are not yet able to make their inferential commitments explicit in the form of conditionals or if-then statements and therefore they cannot be disposed to explicitly acknowledge inferential commitments. These primitive scorekeepers instead have implicit inferential commitments in virtue of performing in
ways that make it appropriate for us logically competent theorists to attribute these inferential commitments to them.

On Brandom’s original account, we can appropriately attribute an inferential commitment to a scorekeeper if that scorekeeper is disposed to update the particular set of assertional commitments and entitlements that she attributes to herself at that time according to that inferential relation (for example attributing commitment to a commitment-preserving relation’s conclusion in response to her attributing commitment to that commitment-preserving relation’s premises). The problem with this, as I argued in Chapter III, is that any time a scorekeeper takes herself to be committed to what she takes to be a defeater for an inferential relation, she would no longer be disposed to update her score according to that inferential relation. Therefore Brandom’s account of the circumstances for attributing inferential commitments conflates an inferential relation’s being defeasible with its being refutable.

The negative consequences of Brandom’s account not allowing for defeaters that are not refuters becomes apparent when we then look at how scorekeepers are obliged (or not obliged) to update score in light of the inferential commitments that Brandom’s account would have us attribute (or not attribute) to them. When a scorekeeper ceases to be committed to an inferential relation because she takes it to be defeated for herself, she also ceases to be obliged to update her interlocutors’ scores according to it—even if the inferential relation is not defeated for those interlocutors. For example, if a scorekeeper correctly takes \{Struck, EMF\} > Unlit to be defeated for herself as a result of her commitment to F Cage, then on Brandom’s account that scorekeeper will cease to be committed to that inferential relation and will then cease to be obliged to remove
entitlement Lights from her interlocutors’ scores even if they are committed to Struck and EMF, but not to FCage.

**Material Inference and the Priority of Propositions**

As I discussed in Chapter IV, one of the challenges in revising Brandom’s account of defeasible reasoning is that it holds that language users must be interpretable as committed to whole propositions standing in particular material inferential relations in order to be interpretable as taking the concepts that compose these propositions as having substitutional-inferential roles (contents). For example, we cannot interpret a scorekeeper as committed to the entitlement-preserving relation *Tweety is a bird > Tweety can fly* in virtue of our being able to interpret that scorekeeper as committed to *x is a bird > x can fly*, because on Brandom’s account we interpret a scorekeeper’s commitments regarding the predicates *is a bird* and *can fly* on the basis of how we interpret that scorekeeper to be taking those predicates to contribute to the inferential roles of whole propositions, like ‘Tweety is a bird’ and ‘Tweety can fly’.

This places significant constraints on the ways we can plausibly revise Brandom’s account of the circumstances for attributing inferential commitments and makes Brandom’s position incompatible with many (if not most) of the accounts of defeasible reasoning and inference in the AI literature. The revised account of defeasible inference will therefore have to provide a way for us to understand defeasible inferential relations like the one leading from ‘The match is struck’ to ‘The match will light’ without appealing to matches generally. Very roughly, the approach to be adopted here understands the generality behind defeasible inferential relations involving claims about
particular birds and particular matches in terms of the ranges of collateral premises we anticipate playing a role in the conversation.

Before I begin presenting my proposed revisions, I warn the reader that the proposed revisions are rather complex and are in some ways not especially conservative. However, I think that the resulting position will maintain much of the general spirit behind Brandom’s account better than the positions that would result if we tried to incorporate more popular conceptions of defeasible reasoning in Brandom’s original account. If the revisions that I propose give rise to a much more complex position than Brandom’s original, then this may be largely part due to the fact that Brandom’s own account of the circumstances for attributing inferential commitments is not worked out very precisely. (For example, Brandom glosses over what he means when he says that a scorekeeper is disposed to update score in such and such a manner.)

**PART 2: Accomodating Inferential Holism**

Clarifying the Sort of Holism to be Accommodated

Because inferential holism is so central to Brandom’s account of inferential commitment (and because I think it is right), it is not a viable option to revise Brandom’s account in such a way that it will not accommodate inferential holism. Unfortunately, it is very difficult to adequately accommodate inferential holism. Inferential holism is the central challenge to be overcome here, and much of the resulting account is the way it is
in order to accommodate it. However, as I will now show, we can aim to accommodate inferential holism generally without trying to accommodate the particular (and particularist) version of it that Brandom adopts.

To start, it is useful to distinguish the following theses:

- Score Updating Holism
- Inferential Holism Generally
- Brandom’s Particularist Inferential Holism

I describe *Score Updating Holism* as roughly the thesis that a scorekeeper’s score updating dispositions are wholly indeterminate unless and until we specify (or at least make an assumption about) the entire set of collateral premises that the scorekeeper is using. As a result we cannot first identify a scorekeeper’s general set of scorekeeping dispositions with regards to individual claims, and then use this information to derive how the scorekeeper would update score were she to use different combinations of these claims as collateral premises (as the atomist would try to do).

I will treat *Inferential Holism Generally* as Score Updating Holism combined with the claim that a scorekeeper’s inferential commitments depend in some unspecified way on her score updating practices. This additional claim may seem obvious to anyone who already accepts a *use* theory of meaning, but it is nonetheless significant in that it rules out accounts of inferential commitment that are more representationalist in nature. For example, it rules out the position that someone is committed to a particular inferential relation in virtue of their having some proposition expressing that inferential relation
tokened in their belief box. I will often just refer to Inferential Holism Generally as ‘inferential holism’.

Brandom’s Particularist Inferential Holism can be understood as one way of fleshing out inferential holism generally. It makes the further claim that a scorekeeper’s inferential commitments at a particular scorekeeping stage are determined by how that scorekeeper is disposed to update the particular score set that she takes herself to have at that stage.

With these distinctions in place we can aim to accommodate Brandom’s commitment to inferential holism generally without attempting to accommodate Brandom’s particularist inferential holism—thereby allowing us to try to fill in the details of inferential holism differently than Brandom does.

One of the main ways that I aim to accommodate inferential holism (in general) in my revised account of the circumstances for attributing inferential commitments is by using the ScoreUpdate function (introduced in Chapter III) to generate the raw data we are to use to justify our attributions. We are entitled to attribute inferential commitments in virtue of the relationships between ScoreUpdate’s inputs and outputs. By using ScoreUpdate functions in this way I make no attempt to treat individual claims (or small sets of them) as having inferential consequences on their own (independently of a background set of collateral premises). In this regard my revised account will differ little from Brandom’s original. Where my account will differ from Brandom’s is in how it will

65 Recall that the ScoreUpdate function has three arguments: a scorekeeper, a score set, and a set of changes made to the score set. Its value (or output) is another score set. For example, ScoreUpdate(S, A, +C{p}) outputs the score set that results from scorekeeper S adding and removing commitments and entitlements to score set A in response to commitment p being added to A.
have us interpret inferential commitments from the raw data. Once again, I emphasize that we are NOT trying to figure out how a scorekeeper would update score in a wide variety of contexts. We are instead assuming that we already know this and are trying to figure out how to read off a scorekeeper’s inferential commitments from these data.

**Challenges in Indentifying Appropriate Reference Score Set(s)**

A scorekeeper’s inferential commitments cannot be used explain or predict how a particular scorekeeper would or should update score in response to any possible change being made to any possible set of collateral premises. Instead, the set of inferential relations that a scorekeeper is committed to is relative to the particular score set (or set of score sets) that the scorekeeper is using as collateral premises. As I explained in Chapter III, we can call the particular score set(s) that a scorekeeper’s inferential commitments are dependent upon reference score set(s). I take it that a large part of fleshing out an acceptable version of inferential holism consists of identifying an appropriate reference score set (or score sets).

In Chapter III, I argued that the key problem with Brandom’s original account (and his particularist inferential holism) is that it makes the reference score set consist only of the particular set of commitments and entitlements that the scorekeeper being discussed takes herself to have at the time in question. If an instance of reasoning requires that a scorekeeper take an inferential relation to be defeated for herself then she will not update score according to that inferential relation, and therefore cannot be appropriately interpreted as implicitly committed to it. This suggests that the reference score set determining a scorekeeper’s inferential commitments must often be distinct from the
score set that she takes herself to have (or at least when she takes an inferential relation to be defeated for herself.)

However, as I discussed at the end of Chapter III, any suggestion to the effect that the reference score set should be some score set that the scorekeeper could (but does not actually) take herself to have needs to be accompanied by an account of why we should use the suggested score set as the reference score set rather than any other score set that the scorekeeper could take herself to have. It is not enough to appeal to how a scorekeeper would update score were she to use a different set of collateral premises. For any inferential relation $IR_x$, we can fabricate some score set $A$ such that were $S$ to use $A$ as collateral premises then $S$ would be committed to $IR_x$. Similarly, For any inferential relation $IR_x$ we can fabricate some score set $B$ such that were $S$ to use $B$ as collateral premises then $S$ would not be committed to $IR_x$. Therefore, if we want claims of the form ‘$S$ is committed to inferential relation $IRA$’ to carry any significance, then we must restrict the range of possible score sets that may be used as reference scoresets.

Towards a Solution: Appeals to Likely Score Sets and Propositions

One of the worries that philosophers often have about holism is that the inferential consequences (and hence meaning) of a potential commitment may change from one point in a conversation to the next. This is a legitimate worry. Making use of the language-game analogy, we don’t mind different games having different rules from each other (perhaps even using largely the same playing pieces in the different games), but we don’t want the rules of any particular game we are engaged in to change every time a new move is made. Likewise, there is very little reason to care about what inferential role a
claim stands in if this role frequently changes (in relevant ways) throughout a conversation. Brandom’s account suffers from this to some extent, which is why I say that it is particularist.

While we don’t want inferential relations to readily switch from being materially good to being materially bad and back to being good again within a conversation, there is no problem with a defeasible inferential relation starting off undefeated for someone, becoming defeated, then becoming undefeated again within a conversation. As a result, an inferential relation’s status with regards to its being defeated or not, does not need to be nearly as stable as its status with regards to being materially good or not.

The revised account that I am developing here tries to stabilize meanings (and inferential commitments) by relativizing them to conversations (or ranges of conversations) rather than to the particular score sets that are attributed or acknowledged at particular stages in conversations. *Once I am able to show how a scorekeeper can be stably committed to inferential relations being materially good for a whole conversation, I will then be able to show how that scorekeeper can remain committed to an inferential relation even while she takes it to be defeated for herself at a particular point in the conversation.*

Suppose that scorekeepers are reasonably reliable at foreseeing the range of score sets that they are likely to attribute to themselves and their interlocutors in a particular conversation or range of conversations. Also suppose that scorekeepers are reasonably reliable at picking out the range of changes that are likely to be made to these score sets throughout the course of the conversation—anticipating a range of commitments and entitlements that they and their interlocutors are likely to gain or lose and which are likely
to make a difference to the giving and asking for reasons in the conversation. Let us say that these are what the scorekeeper in question takes to be the likely score sets and likely propositions.

For present purposes we need not care why scorekeepers take some score sets and propositions to be likely rather than some others. These details might be usefully imported from an already well developed theory about conversation, such as relevance theory. For now, all that matters is that scorekeepers will take only a small portion of all possible score sets and propositions to be likely, and that for each likely proposition there will be some likely score set that contains it and some likely score set that does not.

While my revised account will appeal to the sets of score sets and propositions that the scorekeeper takes to be likely, it would be rather misleading to say that the account I am proposing attempts to identify which inferential commitments the scorekeeper takes to be likely used in the conversation. Doing so mistakenly suggests that we have some antecedent means of identifying inferential relations that the scorekeeper is committed to but does not take to be likely. If we had some other effective means of identifying a scorekeeper’s inferential commitments then obviously we would not need the account being presented here.

Someone may wish to respond saying that even if we do not presently have the means to identify a scorekeeper’s other inferential commitments (the ones the scorekeeper does not take to be likely), we should presume that these inferential

---

66 The reason why scorekeepers must be reliable at anticipating the score sets that they will attribute and the changes that are likely to be made to these score sets will be explained below in Part 4.

commitments are nonetheless implicit in how a scorekeeper updates scores and give rise to the surface scorekeeping regularities that we appeal to when attributing inferential commitments. In response to this concern, I would remind the objector that by adopting inferential holism we give up on assuming that there is some underlying rule structure that can explain or rationalize why a scorekeeper is disposed to update score in the way that she is for each possible change that could be made to each possible score set. We cannot do as the atomist does (or attempts to do), identifying the inferential relations that various propositions stand in on their own and then go on to derive how any given score set would or should be updated in virtue of the inferential relations that each of its members stands in. This is not possible because propositions can only stand in inferential relations against a large background of collateral premises, and we cannot in principle identify all and only those collateral premises that a scorekeeper’s having or not having a particular inferential commitment is contingent on.

As far as the position being developed here is concerned, primitive scorekeepers (those who cannot make their inferential commitments explicit) have no inferential commitments beyond the ones that can be appropriately attributed according to the account to be presented. Furthermore, because scorekeepers are only committed to inferential relations between claims or sets of claims that the scorekeeper takes to be likely, we cannot assess what score sets or propositions a scorekeeper takes to be likely on the basis of that scorekeeper’s inferential commitments.
PART 3: The Revised Account of the Appropriate Circumstances for Attributing Implicit Commitments to Indefeasible Inferential Relations

My revised account of the appropriate circumstances for attributing commitment-preserving and incompatibility relations (the two indefeasible types of inferential relations) will be significantly simpler than my revised account for attributing entitlement-preserving relations (the indefeasible type of inferential relation). When we are trying to identify the conditions for attributing to a scorekeeper various commitments to commitment-preserving and incompatibility relations, I propose that instead of using one particular score set as the reference score set, we should instead use a set of reference score sets. More specifically, I propose that we use all and only those score sets that the scorekeeper takes to be likely. The idea behind this is that scorekeepers do not typically accept an indefeasible inferential relation if they anticipate its being violated when applied to some score sets. Indefeasible inferential relations are, after all, supposed to be free of exceptions.

However, scorekeepers are not implicitly committed to an indefeasible inferential relation’s holding for every possible score set (where each score set acts as possible set of collateral premises). Otherwise, scorekeepers would not be committed to any indefeasible inferential relations. This is because inferential holism has us presume that for any inferential relation there is always some possible score set (perhaps one that nobody actually holds, or one which we cannot presently identify) which the scorekeeper is disposed to update in some way that violates that inferential relation (for example, not attributing commitment to the conclusion of a commitment-preserving relation in
response to attributing commitment to its premises). As discussed at length in Chapter III, this holds true for every inferential relation regardless of whether or not they are defeasible.

In order to accommodate inferential holism, scorekeepers must only be implicitly committed to indefeasible inferential relations holding for all *likely* score sets, rather than all *possible* score sets. The position that results is that a scorekeeper is committed to a particular indefeasible inferential relation if (among other things) that scorekeeper is disposed to update every score set that she takes to be likely according to some pattern specified by the inferential relation (for example adding commitment to a commitment-preserving relation’s conclusion in response to adding commitment to that commitment-preserving relation’s premises).

**Appropriate Circumstances for Attributing Commitments to Commitment-Preserving Relations:**

For any set of propositions \(X\) and any proposition \(y\)
- Scorekeeper \(S\) is implicitly committed to \(X \rightarrow y\) IF
  - \(S\) takes \(X\) to be likely
  - \(S\) takes \(y\) to be likely
- For all score sets \(A\), such that \(S\) takes \(A\) to be likely
  - \(C\{y\} \subseteq \text{ScoreUpdate}(S, A, +C\{X\})\)
  - \(C\{X\} \not\subseteq \text{ScoreUpdate}(S, A, -C\{y\})\)
  - \(E\{y\} \subseteq \text{ScoreUpdate}(S, A, +E\{X\})\)
  - \(E\{X\} \not\subseteq \text{ScoreUpdate}(S, A, -E\{y\})\)
Appropriate Circumstances for Attributing Commitments to Incompatibility

Relations:

For any sets of propositions \( X \) and \( Y \)

Scorekeeper \( S \) is implicitly committed to \( X \upharpoonright Y \) IF

\( S \) takes \( X \) to be likely

\( S \) takes \( Y \) to be likely

For all score sets \( A \), such that \( S \) takes \( A \) to be likely

\[ E\{Y\} \not\subseteq \text{ScoreUpdate}(S, A, +C\{X\}) \]

\[ E\{X\} \not\subseteq \text{ScoreUpdate}(S, A, +C\{Y\}) \]

The account just given does not require that either of the sets of claims involved in the inferential relations must be (or must not be) a subset of the score set being updated (Score set \( A \)). This may seem to lead to one of the problems faced by AGM (as discussed in an earlier chapter), where if we AGM-revise a set of beliefs \( B \) by \( p \), then among \( p \)’s inferential consequences would be any claim that was already a member of \( B \) (and which was not removed in the revision process). However, the present account avoids this by (1) requiring that the inferential relations that a scorekeeper is committed to only relate sets of claims that the scorekeeper takes to be likely and (2) by appealing to how the scorekeeper would update each score set that she takes to be likely, rather than just the particular score set that she takes herself to have at that time. There will be no nonempty sets of likely propositions that will be subsets of every likely score set because propositions are likely only if they are members of some of the likely score sets and not members of other likely score sets. Therefore, there is no danger of a claim \( p \)’s having some claim \( q \) as an inferential consequence simply in virtue of \( q \) being a member of every likely score set.
PART 4: The Revised Account of Appropriate Circumstances for
Attributing Implicit Commitments to Defeasible Inferential Relations

Motivating REF: Using a Shared Common Ground as the Reference Score Set

Because scorekeepers themselves recognize that there may be points in the conversation where a defeasible inferential relation’s conclusion should not be attributed when its premises are, there must be some appropriate conditions for attributing commitments to defeasible inferential relations that do not require that the scorekeeper would update each likely score set according to some specified pattern (for example attributing entitlement to the inferential relations’ conclusion in response to attributing commitment and entitlement to its premises). Simply put, while scorekeepers are implicitly committed to indefeasible inferential relations being strict and free of exceptions, they are not committed to defeasible inferential relations being this way.

Instead of using all the likely score sets as the reference point for attributing commitments to defeasible inferential relations (as I proposed that we do for indefeasible inferential relations), I propose that we use a single reference score set that will act as a sort of shared hypothetical starting place for the conversation. Once we have identified the appropriate reference score set then we will proceed in largely the same way that we would using Brandom’s original account: we will identify what a scorekeeper takes to be a set of claim X’s inferential consequences by looking at how that scorekeeper would update the references score set in response to commitment and entitlement to X being added to it.
The sort of hypothetical shared starting place for the conversation that I have in mind is not necessarily (and not typically) the set of commitments and entitlements that each (or any) interlocutor actually brings into the conversation, as if each interlocutor came into the conversation with the same set of commitments and entitlements. It rather purports to be a point of reference that a scorekeeper uses (and takes her interlocutor to be using) to make sense of the different assertions likely to be made in a conversation.

This might be usefully understood in terms of a sort of card game, where the different commitments that different interlocutors bring into the conversation can be thought of as different hands of cards dealt to each person. Players make moves (make assertions) in the game by publicly playing one of the cards in their hand (thereby making their commitment explicit). Each player keeps score on every other player on the basis of the cards actually played. Therefore, whether or not a player takes an interlocutor to be entitled to play a particular card will depend upon what other cards have already been played, but not upon any of unplayed cards (still held in any of the players’ hands). At the start of the game (conversation), each player has the same public score even though each will have been dealt a different hand of cards (commitments). It is with reference to this purportedly shared starting point of the game that players each evaluate how the playing of a certain card or combination of cards affects score.

One of the central ideas motivating the approach being adopted is the commonly recognized principle that when reasoners are giving arguments they should avoid using as premises any claims that their interlocutors are unlikely to accept (whether the interlocutors deny them or are agnostic about them). If a reasoner wants to use a premise $p$ that he suspects an interlocutor does not accept then (1) the reasoner will need to make
I take it that these claims are themselves largely uncontentious, as we often tell them (or claims similar to them) to our undergraduates when teaching them how to give arguments.

The principle that we should not argue from contentious premises is probably most obvious when dealing with premises that are explicitly stated, but it is all the more crucial to avoid using any contentious collateral premises that remain implicit in the background. This is simply because the collateral premises implicit in the background largely determine the scorekeeping significance of the premises that are explicitly stated. Reasoners who are implicitly using relevantly different sets of collateral premises will not be in a position to argue for or against claims because they would likely not agree on what follows from the premises that they would each assert. Because we are dealing with material inference rather than formal inference, reasoners would not even be on the same page about what some assertion of a claim is supposed to be a reason or refuter for. This makes it difficult for them to even identify disagreement. Therefore, if a scorekeeper wants to use a premise that she anticipates her interlocutors may not accept, then for the sake of clarity and effective conversation she should assert it as a co-premise, rather than leave it as an implicit collateral premise.

---

68 I say ‘may’ because, depending on the context and the claim, the scorekeeper’s interlocutors can become entitled to claims without having reasons for those claims. For example, the scorekeeper may be entitled to her asserting ‘The match is struck’ because she is reliable at making such observations. So even though the scorekeeper’s interlocutors might not have initially been committed or entitled to the claim ‘The match is struck’ they can become entitled to it by overhearing the scorekeeper assert it and by taking the scorekeeper to be reliable.
The norm that reasoners should try to identify and then avoid arguing from contentious claims (either implicitly or explicitly) is not merely optional for reasoners to understand one another in conversation. It is necessary, playing much the same role in my account as the principle of charity does in Davidson’s. Both are ways of trying to overcome/accommodate some of the challenges raised by holism.\textsuperscript{69} But rather than have us assume that we are on the same page as our interlocutors, by having us go into conversations defeasibly presuming that our interlocutors have the same beliefs that we do, my account forces us to anticipate likely differences between what we accept and what our interlocutors do—thereby increasing clarity by minimizing misunderstanding. If we take a claim to be contentious then we should not use it as a collateral premise \textit{even if it is true}, unless and until we make it explicit and our interlocutors accept it as well.

The danger of doing otherwise is not the common sort of misunderstanding that we can commonly rectify through the course of a conversation. For any such misunderstanding (about what claims an assertion is supposed to be a reason or refuter for) there are likely infinitely many differences in collateral premises that could give rise to it. Unless reasoners were generally reliable at identifying likely contentious claims and avoiding using them as collateral premises then they could not hope to successfully rectify these sorts of misunderstandings. If reasoners did not generally avoid using

\textsuperscript{69} See Davidson’s “Radical Interpretation” (137). While there are indeed some important similarities between Brandom’s and Davidson’s accounts, there are also important differences both in regards to the details of their positions and in regards to the questions that they are trying to answer. For example, Davidson is much more committed to naturalism than Brandom is, and therefore it is crucial to Davidson’s, but not Brandom’s account, that a theory of meaning be empirically testable. Also, while Davidson tries to explain interpretation between users of the same language in terms of what a theorist does in radical interpretation, Brandom would try to explain radical interpretation in terms of what speakers of the same language do unreflectively.
contentious claims as collateral premises (or were generally unreliable at it) then a complete breakdown of communication would occur and they would not even be able to misunderstand one another.

**Defining REF:**

If claims stood in inferential relations independently of any collateral premises then we could perhaps use an empty set of commitments and entitlements as the shared starting place for the conversation. But inferential holism tells us that claims only have inferential consequences relative to very large sets of collateral premises, and therefore we must identify a large set of commitments and entitlements that can function as a scorekeeper’s perspective on the conversation’s shared starting place. We will then use this large set of commitments and entitlements to identify the defeasible inferential relations that that scorekeeper is committed to.

Here is what I propose: let the hypothetical shared starting place for some conversation consist of the intersection of all the score sets that are likely to be used in that conversation. It is crucial that we only consider the intersection of all likely score sets, rather than all possible score sets because the intersection of all possible score sets would be empty (or at least much too small) and therefore could not function as the reference score set. After all, for any proposition, there is some possible score set which does not include commitment or entitlement to it.

---

70 More accurately, it will be the score set whose set of commitments consists of the intersection of all the sets of commitments belonging to all likely score sets, and whose set of entitlements consists of the intersection of all the sets of entitlements belonging to all the likely score sets.
The set of propositions in any of the likely score sets will not be limited to just those propositions that are likely. Most of a reasoner’s commitments and entitlements will not change throughout a conversation. Furthermore, and here is the crucial bit, most of the claims that are not likely to be used in the conversation, but which are members of at least some of the likely scoresets will also be members of every other likely scoreset. Those engaged in a conversation will be committed and entitled to many of the same propositions throughout the conversation and the scorekeeping carried out in the conversation will largely depend on the scorekeepers having these shared commitments and entitlements. However, the propositions that interlocutors share commitment or entitlement to cannot be considered likely since propositions are only likely if they are members of some likely score sets and not others (because commitment to the likely propositions may be gained or lost in the conversation). The same applies with regards to interlocutors in a conversation not being committed or entitled to many of the same propositions.

It is this shared set of non-likely commitments and entitlements that I will use as the reference score set for determining a scorekeeper’s commitments to entitlement-preserving inferential relations. Let us call the reference score set $REF$, and define it as follows:

Let $REF =$ The intersection of the sets of commitments belonging to all likely score sets, and the intersection of the sets of entitlements belonging to all likely score sets.

As to be expected, however, the set of score sets that a scorekeeper takes to be likely is often not the same as the set of score sets that are in fact likely. Scorekeepers
might be reliable at anticipating the various score sets they will likely attribute in a conversation, but they are not infallible. Different scorekeepers may have different expectations about what score sets they and their interlocutors are likely to have, even though there is a norm to try and synchronize these expectations. But if we want to assess a scorekeeper’s inferential commitments then we will need to consider what she takes to be likely. REF will therefore be different from the different perspectives of different scorekeepers. We can therefore identify whose perspective on REF we are discussing by indicating it in brackets.

Let $REF(S) =$ The intersection of the sets of commitments belonging to all score sets that $S$ takes to be likely, and the intersection of the sets of entitlements belonging to the score sets that $S$ takes to be likely.

To help understand how using REF as the reference score set helps us, let us consider an example using the following abbreviations:

$Struck =$ The dry, well-made match is struck

$Lights =$ The match lights

$Unlit =$ The match remains unlit

$EMF =$ The match is in a strong electromagnetic field

$FCage =$ The match is in a Faraday cage

$Vacuum =$ The room is evacuated of oxygen

$Oxygen =$ The room has oxygen in it

$WoodEnd =$ The wooden end of the match is being held
$RoomTemp = \text{The match is at room temperature}$

Now let us suppose that the scoresets indicated below are the only ones that the scorekeeper takes to be likely to be used in some conversation about whether the match will light or not. As in Chapter III, I use ‘$C$’ to prefix a set of commitments, and ‘$E$’ to prefix a set of entitlements. For simplicity’s sake, I use ‘$\ldots$’ to indicate that the sets of commitments and entitlements contain many more propositions than I actually list. This is admittedly a bit of a cheat because it effectively says ‘and all the various other claims that we would normally expect to be included’.

Likely Score Sets (according to $S$):

$$Score\ set\ A = C\{\text{Struck, WoodEnd, RoomTemp, Oxygen, $\ldots$}\}$$
$$E\ \{\text{Lights, Struck, Fcage, WoodEnd, RoomTemp, Oxygen, $\ldots$}\}$$

$$Score\ set\ B = C\ \{\text{Struck, EMF, WoodEnd, RoomTemp, $\ldots$}\}$$
$$E\ \{\text{Unlit, EMF, Struck, WoodEnd, RoomTemp, Oxygen, $\ldots$}\}$$

$$Score\ set\ C = C\ \{\text{Struck, Fcage, WoodEnd, RoomTemp, $\ldots$}\}$$
$$E\ \{\text{Lights, EMF, Fcage, Struck, WoodEnd, RoomTemp, $\ldots$}\}$$

$$Score\ set\ D = C\ \{\text{EMF, Vacuum, WoodEnd, RoomTemp, $\ldots$}\}$$
$$E\ \{\text{Unlit, Vacuum, EMF, Fcage, WoodEnd, RoomTemp, $\ldots$}\}$$

$$Score\ set\ E = C\ \{\text{EMF, Fcage, WoodEnd, RoomTemp, $\ldots$}\}$$
$$E\ \{\text{EMF, Fcage, WoodEnd, RoomTemp, $\ldots$}\}$$

Let us also suppose that the set of likely propositions are the ones indicated below.

Likely Propositions (according to $S$):

$$\{\text{Struck, EMF, Fcage, Vacuum, Lights, Oxygen, Unlit}\}$$
If a scorekeeper’s commitments to entitlement-preserving relations were based on any (and only) one of these score sets then she would be committed to entitlement-preserving relations which interlocutors who accept different likely score sets would not (and should not) accept. Such a scorekeeper’s inferential commitments would be insensitive to her having commitments and entitlements that her interlocutors do not, and her interlocutor’s having commitments and entitlements that she does not. In this context, each interlocutor (each using a different set of collateral premises) would make inferences in their giving and asking for reasons that other interlocutors would find objectionable. For example an interlocutor using score set $B$ (which includes commitment and entitlement to $EMF$ but not $FCage$) as collateral premises would take $Struck$ to have different consequences than someone who is using score set $E$ (which includes commitment and entitlement to both $EMF$ and $FCage$). Even worse, the different interlocutors will be prone to argue past one another because they may not realize which collateral premises their interlocutors are making use of and may not be on the same page about what commitments an assertion is supposed to be a reason or refuter for.

By using $REF(S)$ we avoid these problems because each interlocutor only uses as collateral premises those commitments and entitlements that they expect every other interlocutor to hold throughout the conversation. Each of the likely propositions is considered contentious in the conversation and cannot (for the moment) be assumed by any one of the interlocutors when giving and asking for reasons. What, according to the scorekeeper, they may each assume as collateral premises is $REF(S)$, where in the present example
While much of the score updating in the conversation may be contingent on scorekeepers using the claims \textit{WoodEnd} and \textit{RoomTemp} as collateral premises, these claims are not contentious and therefore each interlocutor in the conversation may use them (and may be expected to use them) as collateral premises in their giving and asking for reasons.

\textbf{How we Use REF to Identify a Scorekeeper’s Commitments to Entitlement-Preserving Relations}

Recall from Chapter III that Brandom’s original account of the appropriate circumstances for attributing commitments to entitlement-preserving relations was as follows:\textsuperscript{71}

1) For any set of claims \(X\) and any claim \(y\)
2) Scorekeeper S is implicitly committed to \(X > y\) IF
3) \(A = \text{whatever particular score set S takes herself to have}\)
4) \(+E\{y\} \in \text{ScoreDiff}(A + CE\{X\}, \text{ScoreUpdate}(S, A, +CE\{X\}))\)

On Brandom’s original account, the reference score set is whatever score set the scorekeeper (S in this case) takes herself to have, and we assess what S takes to be the entitlement-preserving consequences of premise set \(X\) by looking at what entitlements S would add to her score set in response to commitment and entitlement to \(X\) being added to it.

\textsuperscript{71} As explained in Chapter III, the function \textit{ScoreDiff} takes two score sets as arguments and outputs a set of changes that must be made to the first score set in order to make it identical to the second.
The key difference between Brandom’s original and my revised account of attributing commitments to entitlement-preserving relations is our using different reference score sets. I also add the additional requirement that the sets of claims in the inferential relations must be likely (or rather taken to be likely). But we cannot simply replace whatever score set scorekeeper $S$ takes herself to have with $REF(S)$. The problem is that $S$ is likely to take $REF(S)$ to be what I call an unstable score set. $S$ takes some score set $A$ to be unstable if and only if $S$ would be disposed to update $A$ even if no changes have been made to $A$. If we uses empty set brackets ‘$\{\}$’ to indicate an empty set of changes that are being made to a score set $A$ then we can use the following definition.

For any score set $A$,

$S$ takes $A$ to be unstable IFF $A \neq ScoreUpdate(S, A, \{\})$

The main reason why a scorekeeper $S$ would take $REF(S)$ to be unstable is because $REF(S)$ will not include entitlement to any claims that $S$ takes to be likely, and if $S$ were to update $REF(S)$—but not in response to any changes being made to it—then $S$ would probably add entitlements to claims that $S$ takes to be likely. Continuing our earlier example above, since $S$ takes the claim $Oxygen$ to be likely then entitlement to $Oxygen$ will be a member of some of the score sets that $S$ takes to be likely but not others. This is why $Oxygen$ is not a member of the intersection of all those sets of entitlements belonging to score sets that $S$ takes to be likely. However, even though entitlement to $Oxygen$ is not a member of $REF(S)$, it is plausible to suppose that $S$, upon attributing $REF(S)$, would be immediately disposed to update it to include entitlement to $Oxygen$. That is, it is plausible that

$+E\{Oxygen\} \in ScoreDiff(REF(S), ScoreUpdate(S, REF(S), \{\}))$
The challenge with using $REF(S)$ as the reference score set becomes apparent when we want to identify the entitlement-preserving consequences that $S$ takes different claims like $Struck$ to have. If $S$ takes $REF(S)$ to be unstable in the way that I have just described then the following would likely hold:

$$+E\{Oxygen\} \in ScoreDiff(REF(S), \ ScoreUpdate(S,REF(S),+CE\{Struck\})$$

This misleadingly suggests that $S$ would add entitlement to $Oxygen$ in response to commitment and entitlement to $Struck$ being added—which would incorrectly lead us to interpret $S$ as being committed to $Struck \succ Oxygen$. The problem is that $S$ would add entitlement to $REF(S)$ even if commitment and entitlement to $Struck$ were not added.

However, even though we should not use $REF(S)$ as the reference score set when identifying the inferential consequences that $S$ takes different claims (or sets of claims) to have, we can use $REF(S)$ to identify what $S$ takes herself and her interlocutors to have default entitlement to. In the previous chapter I showed how we can understand default entitlement in terms of the revised entitlement-preserving relations. For example, if $S$ is committed to $\{\} \succ p$, then $S$ is obliged to take commitment and entitlement to the empty set to be a defeasibly sufficient condition for being entitled to $p$. For $S$ to be disposed to update her reference score set according to $\{\} \succ p$ is for $S$ to be disposed to add entitlement $p$ to her reference score set, but not necessarily in response to any change having been made to her reference score set. Because the inferential relations that a scorekeeper is committed to only relate sets of claims that the scorekeeper takes to be likely, we should understand the empty set of premises involved in default entitlements as
an empty set of likely premises. A background set of collateral premises is still necessary.

We can use REF to play that role as follows:

(1) For any claim \( x \),
(2) \( S \) is committed to \{ \} \( \rightarrow \) \( x \) IF
(3) \( S \) takes \( x \) to be likely
(4) \(+E\{x\} \in \text{ScoreDiff}(\text{ScoreUpdate}(S,REF(S),\{\}))\)

In adopting this account of default entitlement we can interpret \( S \) in our ongoing example as being implicitly committed to \{ \} \( \rightarrow \) Oxygen.

However, if we want to attribute to \( S \) commitments to entitlement-preserving relations that do not have empty premise sets then we need to use a reference score set that \( S \) takes to be stable. I suggest that we use a stabilized version of \( \text{REF}(S) \), which I will call \( \text{REF}^\wedge(S) \), and define as follows:

\[
\text{LET} \quad \text{REF}^\wedge(S) = \text{ScoreUpdate}(S,\text{REF}(S),\{\})
\]

\( \text{REF} \) is the absolute shared reference point for a conversation, in that it purports to be a subset of all interlocutors' scores throughout a conversation. \( \text{REF}^\wedge \), on the other hand, is simply \( \text{REF} \) expanded to include entitlements that interlocutors will have by default (in addition to various other commitments and entitlements that are not taken to be likely). These default entitlements may be lost and reacquired throughout the conversation, but every interlocutor will have them unless and until they do something to lose them.

Once we have identified \( \text{REF}^\wedge(S) \) we can use Brandom's original account of the circumstances for attributing commitments to entitlement-preserving relations, except
that (1) we use $REF^\wedge(S)$ as the reference score set instead of the score set $S$ takes herself to have, and (2) we require that $S$ takes the claims involved in the entitlement-preserving relation to be likely.

**Appropriate Circumstances for Attributing Commitments to Entitlement-Preserving Relations:**

For any set of propositions $X$ and any proposition $y$

ScoreKeeper $S$ is implicitly committed to $X > y$ IF

$S$ takes $X$ to be likely
$S$ takes $y$ to be likely
$+E\{y\} \in ScoreDiff(REF^\wedge(S) + CE\{X\}, ScoreUpdate(S,REF^\wedge(S),+CE\{X\})$

Continuing the example above, let us suppose we want to know what entitlement-preserving consequences $S$ takes the set of claims $\{Struck, EMF\}$ to have. To find this out we need to know what entitlements must be added to the score set $REF^\wedge(S) + CE\{Struck, EMF\}$ to make it equivalent to the score set $ScoreUpdate(S,REF^\wedge(S),+CE\{Struck, EMF\})$. Here $REF^\wedge(S)$ is the score set being updated, and $ScoreUpdate(S,REF^\wedge(S),+CE\{Struck, EMF\})$ is the score set that would result if $S$ updated $REF^\wedge(S)$ in response to commitment and entitlement to $Struck$ and $EMF$ being added.

Let us suppose that we know the following:

$$REF^\wedge(S) = C\{Struck, WoodEnd, RoomTemp, ...\}$$
$$E\{Oxygen, Struck, WoodEnd, RoomTemp, ...\}$$

$$ScoreUpdate(S,REF^\wedge(S), [+ CE\{Struck, EMF\}])$$
= \{Struck, EMF, WoodEnd, RoomTemp, ...\}
E\{Unlit, Oxygen, Struck, EMF, WoodEnd, RoomTemp, ...\}

We now look for those entitlements that are in the updated score set but not already members of the original score set being updated (in this case $REF^<(S)$). Each of these will be different entitlement-preserving consequences that S takes the premises \{Struck, EMF\} to have. In our present example we see that S would add entitlement to Unlit in response to commitment and entitlement to Struck and EMF being added to $REF^<(S)$. Therefore we can appropriately attribute to S a commitment to the entitlement-preserving relation \{Struck, EMF\} $\rightarrow$ Unlit. (Thankfully, we don’t actually need to identify all the commitments and entitlements in different score sets when attributing commitments to entitlement-preserving relations. By using the ScoreDiff function we can gloss over these details and focus only on changes made to score sets.)

**PART 5: Applying the Revised Account to the Match-Lighting Scenario**

We now have the resources available to identify the appropriate circumstances for attributing to a scorekeeper S all the inferential relations involved in the match lighting scenario. Again, we will assume that we can already identify or at least point to the set of score sets and propositions that S takes to be likely to be used in the conversation about the match lighting, and that we have therefore identified $REF(S)$ and $REF^<(S)$. We can also assume that S takes the premise sets and conclusions of each of the inferential relations in the match lighting scenario to be likely.
The inferential relations we need to reconstruct the match lighting scenario are as follows:

\[
\begin{align*}
\text{Lights} & \ // \ Unlit \\
{} & > \ Unlit \\
\{\text{Struck}\} & > \ \text{Lights} \\
\{\text{Struck, EMF}\} & > \ Unlit \\
\{\text{Struck, EMF, FCage}\} & > \ \text{Lights} \\
\{\text{Struck, EMF, FCage, Vacuum}\} & > \ Unlit
\end{align*}
\]

The incompatibility relation \(\text{Lights} // \ Unlit\) was not in Brandom’s original example, but (as we have done previously) we use it so that we do not have to assume formal negation. The default \({}\ > \ Unlit\) was not part of the original example either, but I think that it will be illustrative to include at least one default. What follows are the appropriate conditions for attributing to some scorekeeper \(S\) each of the inferential relations involved in the match lighting scenario.

\(S\) is implicitly committed to \(\text{Lights} // \ Unlit\) IF
For all score sets \(A\), such that \(S\) takes score set \(A\) to be likely
\[E\{\text{Lights}\} \notin \text{ScoreUpdate}(S, \text{ScoreSet}(z), +C\{\text{Unlit}\})\]
\[E\{\text{Unlit}\} \notin \text{ScoreUpdate}(S, \text{ScoreSet}(z), +C\{\text{Lights}\})\]

\(S\) is implicitly committed to \({}\ > \ Unlit\) IF
\[+E\{\text{Unlit}\} \in \text{ScoreDiff}(\ \text{REF}(S), \text{REF}^\wedge(S))\]

\(S\) is implicitly committed to \(\{\text{Struck}\} > \ \text{Lights}\) IF
\[+E\{\text{Lights}\} \in \text{ScoreDiff}(\ \text{REF}^\wedge(S) + CE\{\text{Struck}\}, \text{ScoreUpdate}(S, \text{REF}^\wedge(S), +CE\{\text{Struck}\}))\]

\(S\) is implicitly committed to \(\{\text{Struck, EMF}\} > \ \text{Unlit}\) IF
\[+E\{\text{Unlit}\} \in \text{ScoreDiff}(\ \text{REF}^\wedge(S) + CE\{\text{Struck, EMF}\}, \text{ScoreUpdate}(S, \text{REF}^\wedge(S), +CE\{\text{Struck, EMF}\}))\]

\(S\) is implicitly committed to \(\{\text{Struck, EMF, FCage}\} > \ \text{Lights}\) IF
\[+E\{\text{Lights}\} \in \text{ScoreDiff}(\ \text{REF}^\wedge(S) + CE\{\text{Struck, EMF, FCage}\}, \text{ScoreUpdate}(S, \text{REF}^\wedge(S), +CE\{\text{Struck, EMF, FCage}\}))\]
PART 6: How the Revised Account Allows for Defeaters that are not Refuters

In this section, I will show that the revised account of the circumstances for attributing inferential commitments will allow for defeaters that are not refuters. I will first explain how we are to understand refuters on the revised account and then show how they are different from defeaters.

Refuters

On Brandom’s original account, we saw that a scorekeeper may implicitly change her set of inferential commitments as a result of her changing the score set that she takes herself to have. On the revised account, a scorekeeper changes the inferential commitments that she is committed to for some conversation as a result of her changing
the set of score sets or the set of propositions that she takes to be likely to be used in that
conversation.\footnote{A scorekeeper could, of course, change her inferential commitments by arbitrarily
changing the manner in which she would update different score sets. I have been
idealizing scorekeeping practices throughout this dissertation by assuming that
scorekeepers do not typically change their score updating practices except as a result of
their coming to attribute different commitments and entitlements. If these sorts of
arbitrary changes happened too regularly, then miscommunication would be the norm.}

Revised account of refuters:

\[
\text{IF (IF } S \text{ were to take } p \text{ to be likely THEN the conditions for attributing inferential}
\text{commitment } IRa \text{ to } S \text{ would not be satisfied) THEN } S \text{ takes } p \text{ to be a refuter for}
\neg IRa. \footnote{I am hesitant to say that the conditional embedded in the antecedent is also a necessary
condition for } S \text{'s taking } p \text{ to be a refuter for } IRa, \text{ because someone capable of using
logical vocabulary may have different ways of taking } p \text{ to be a refuter for } IRa \text{ (for example, by asserting } p \implies \neg IRa)
scorekeeper comes to treat as likely would act as a relevant counter example to the indefeasible inferential relation.

A similar story can be told for defeasible inferential relations. If a scorekeeper comes to take some score set to be likely that she did not previously take to be likely then the intersection of all the score sets that she takes to be likely may change, thus opening the possibility of her changing her perspective on \( \text{REF} \) and \( \text{REF}^\land \). Because a scorekeeper’s commitment to defeasible inferential relations is largely determined by her perspective on \( \text{REF} \) and \( \text{REF}^\land \), a scorekeeper could implicitly revise her commitments to defeasible inferential relations by revising the set of score sets that she takes to be likely.

For example, suppose that \( \text{Toy} = ‘\text{The match is a toy}’ \), and that scorekeeper S in our match lighting example does not initially take \( \text{Toy} \) to be likely. Unexpectedly to S, one of her interlocutors makes \( \text{Toy} \) likely by asserting it. When S comes to take \( \text{Toy} \) to be likely she also comes to take a whole new range of score sets to be likely, and as a result her perspective on \( \text{REF} \) changes. Whether or not the scorekeeper remains committed to \( \text{Struck} \succ \text{Lights} \) will largely depend upon how S would update her new perspective on \( \text{REF} \), thereby obtaining her new perspective on \( \text{REF}^\land \). If S takes herself and her interlocutors to be able to assume by default that the match is not a toy then she will probably just treat \( \text{Toy} \) as a another defeater for \( \text{Struck} \succ \text{Lights} \) (just as she does for \( \text{EMF} \)). Alternatively, S’s new perspective on \( \text{REF}^\land \) may change sufficiently such that she no longer accepts \( \text{Struck} \succ \text{Lights} \), in which case we can interpret S as taking \( \text{Toy} \) to be a refuter for \( \text{Struck} \succ \text{Lights} \).
Defeaters that are not Refuters

The argument from Chapter III that Brandom’s account does not allow for defeaters that are not refuters was as follows:

Premise 1:

IF(IF S were to take herself to be committed to \( p \) THEN the conditions for attributing inferential commitment \( IRa \) to S would not be satisfied) THEN S takes \( p \) to be a refuter for \( IRa \)

Premise 2:

IF S takes \( p \) to be a defeater for \( IRa \) THEN (IF S were to take herself to be committed to \( p \) THEN S would not take the addition or removal of \( IRa \)’s premises to or from her score set to have any scorekeeping significance with regards to \( IRa \)’s conclusion.)

Premise 3:

IF S would not take the addition or removal of \( IRa \)’s premises to or from her score set to have any scorekeeping significance with regards to \( IRa \)’s conclusion THEN the conditions for attributing inferential commitment \( IRa \) to S would not be satisfied.
Intermediate Conclusion (From 2 and 3):

IF S takes \( p \) to be a defeater for \( IRa \) THEN (IF S were to take herself to be committed to \( p \) THEN the conditions for attributing inferential commitment \( IRa \) to S would not be satisfied)

Conclusion (From 1 and 4):

IF S takes \( p \) to be a defeater for \( IRa \) THEN S takes \( p \) to be a refuter for \( IRa \).

I assumed (and continue to assume) that premises 1 and 2 are uncontentious.

Premise 1 can be understood as a special instance of the revised account of refuters given above where S comes to take \( p \) to be likely as a result of her unexpectedly becoming committed to \( p \). I spent most of Chapter III arguing that Premise 3 is a consequence of Brandom’s account of the appropriate circumstances for attributing inferential commitments. What I will now briefly show is that*Premise 3 is not a consequence of the revised account of the appropriate circumstances for attributing inferential commitments.*

One of the central features of the revised account of the conditions for attributing commitments to entitlement-preserving relations is that the reference score set remains largely unchanged throughout the conversation even though the scorekeeper (to whom we are attributing the inferential commitments) attributes different sets of commitments and entitlements to herself and her interlocutors at different points in the conversation. Therefore, even if at some point in the conversation the scorekeeper takes herself (or her interlocutors) to have some score set that she is not disposed to update according to some entitlement-preserving relation, the scorekeeper can still be appropriately interpreted as
committed to that inferential relation if she would update the *reference score set* according to it.

For example, suppose S is committed to $FCage$ and as a result is not disposed to update her score set according to \{Struck, EMF\} $\rightarrow$ Unlit (ie. S is not disposed to add entitlement to Unlit and remove entitlement to Lights in response to commitment and entitlement to Struck and EMF being added to her score set). On Brandom’s original account, where the reference score set is the scorekeeper’s own particular score set, we should not attribute to S a commitment to \{Struck, EMF\} $\rightarrow$ Unlit. However, on the revised account it does not matter how S would update her own particular score set, but only how she would update $REF^+(S)$. Therefore, on the revised account S can take an entitlement-preserving relation to be defeated for herself (or her interlocutors) while continuing to be implicitly committed to it.
Summary and Concluding Remarks

This dissertation began with a general discussion of Brandom’s view that a sentence’s propositional content consists of the role that sentence plays as premise or conclusion in three different types of material inferential relations: commitment-preserving, entitlement-preserving, and incompatibility relations. I then described how Brandom tries to ground each of these inferential relations in unreflective, and pre-logical scorekeeping practices. The role of logical vocabulary, Brandom claims, is to make explicit the material inferential commitments that we are implicitly committed to in practice. Once we make them explicit, we can rationally reflect on (give and ask for reasons about) the correctness of our material inferences, and hence rationally reflect on the appropriateness of our non-logical concepts which incorporate these inferences. If the appropriate circumstances for using some concept do not actually warrant the consequences of using that concept, then the concept incorporates a materially bad inference.

On Brandom’s account, the circumstances for appropriately attributing to a scorekeeper a commitment to any of the three types of inferential relations are that scorekeeper’s being disposed to update score (adding and removing commitments and entitlements) in particular manners. The consequences of attributing commitments to the different types of inferential relations are the scorekeeper’s being obliged to update score in particular manners. The scorekeeping vocabulary that Brandom uses to describe his position acts as a sort of meta-vocabulary for articulating (among other things) these circumstances and consequences.
In the first half of this dissertation I argued that Brandom’s three types of material inferential relations themselves incorporate materially bad inferences when applied to common instances of defeasible reasoning (including Brandom’s own match-lighting example). In Chapter II, I argued that Brandom’s account of the consequences for attributing inferential commitments (most clearly seen in his scorekeeping algorithm) do not allow entitlement-recovery, such that a scorekeeper would be obliged to update score incorrectly in reasoning scenarios where one gains, loses, and then regains entitlement to a claim in response to one’s consecutively acquiring new commitments and entitlements. The problem of entitlement-recovery is independent of any account of the appropriate circumstances for attributing inferential commitments because, on Brandom’s account of the consequences, there simply is no combination of inferential commitments that would oblige a scorekeeper to update score correctly. This was because the only type of inferential relation that could allow claims to defeat another inferential relation is itself indefeasible. This is the first of the two ways that Brandom’s three types of inferential relations together incorporate materially bad inferences.

In Chapter V, I proposed modifications to Brandom’s account of the consequences of attributing inferential commitments so that it could allow entitlement-recovery. This primarily involved adding another way that the defeasible inferential relations in Brandom’s account—the entitlement-preserving relations—could be defeated (and then modifying the scorekeeping algorithm accordingly). Drawing on a version of the principle of specificity, I proposed that the premises of an entitlement-preserving relation IRa can be a defeater for another entitlement-preserving relation IRb if IRa’s and IRb’s conclusions are incompatible, and IRa’s premise set is not a proper subset of IRb’s
premise set. This not only made it possible for the premises of entitlement-preserving relations to defeat some other entitlement-preserving relations, it made it possible for any entitlement-preserving relation to be defeated by the premises of yet another entitlement-preserving relation.

In Chapter III, I presented the second of the two ways that Brandom’s three inferential relations incorporate materially bad inferences. I argued that Brandom’s account of the appropriate *circumstances* for attributing inferential relations does not allow for inferential relations to have defeaters that are not refuters. On Brandom’s account a scorekeeper is committed to a particular inferential relation if she is disposed to update her own particular score according to it. As we saw, the problem with this was that if a scorekeeper took herself to be committed to a defeater for an inferential relation, then that scorekeeper would not be disposed to update score according to that inferential relation and would not count as committed to it. This is a problem because, on Brandom’s account, a scorekeeper is only obliged to update the scores of herself and her interlocutors according to her own set of inferential commitments. A scorekeeper disposed to update her own score correctly when committed to a defeater for some inferential relation would be obliged to incorrectly update the scores of any interlocutors who are not committed to that defeater (or any other defeater for that inferential relation).

The solution to the problem of defeaters that are not refuters was just discussed above, and consists of attributing inferential commitments to a scorekeeper on the basis of how that scorekeeper is disposed to update likely score sets and the intersection of likely score sets rather than how she is disposed to update her own particular score set. The intersection of all score sets that the scorekeeper takes to be likely represents that
scorekeeper’s perspective on the shared starting place for whatever conversation she is engaged in. It is relative to this shared starting place that the scorekeeper assesses what a set of claim’s entitlement-preserving consequences are (even though it will probably not be identical with the score that she takes herself to have at any point in the conversation).

The revised accounts of the circumstances for and consequences of attributing inferential commitments do not just solve the problems of entitlement-recovery and the problem of defeaters that are not refuters. They also begin to provide a new way for us to understand the role of defeasible inference within social and holistic linguistic practices. On the revised accounts, the primary importance of defeasible inferential relations is not that they allow reasoners to make rough generalizations about classes of objects, events, or whatever else might be the topic of conversation. Rather, *the primary importance of defeasible inferential relations is that in the context of inferential holism, they allow reasoners to successfully give and ask for reasons despite their only have rough expectations about which commitments each of their interlocutors are bringing into the conversation.*
Works Cited:


Gärdenfors, Peter. “The Dynamics of Belief as a Basis for Logic.”


Poole, D.L. “On the Comparison of Theories: Preferring the Most Specific Explanation.”


Rosenkranz, Sven. “Farewell to Objectivity: A Critique of Brandom.”


Whiting, Daniel J. “Meaning Holism and De Re Ascription.”