Defeasibility in Epistemology

Dissertation Summary

This dissertation is an exercise in what we might call "nontraditional formal epistemology". It's both common and natural to define formal epistemology in opposition to its more traditional cousin: Where traditional epistemology approaches (normative) questions relating to belief, knowledge, and reasoning relying on the classical method of conceptual analysis, formal epistemology approaches these same questions drawing on tools from math and logic. In principle, various formal tools could be used, but, in practice, formal epistemologists usually draw on either the Bayesian framework—which combines probability theory and inductive logic—or epistemic logic. The use of these standard tools to answer epistemological questions can, then, be called "traditional formal epistemology", and what's going on in this dissertation can be thought in opposition to it. In any event, it tackles epistemological questions drawing on a completely different formal framework, namely, that of logics for defeasible reasoning.

This framework originated in the field of artificial intelligence in response to the challenge to represent the information that would let a machine exhibit intelligent behavior. Efforts to meet this challenge quickly made it clear that ordinary logic is utterly inadequate for the task, since much of the information in question takes the form of defeasible generalizations. Thus, the statements "Birds fly" and "Things that look red are red" appear to express sensible principles of reasoning—principles we seem to constantly rely on in our everyday lives—even though they allow for exceptions. Defeasible logics then are, roughly, logics of such defeasible generalizations, and the thesis of this dissertation is that they can be of great help in answering important normative questions in epistemology. The thesis is supported by developing a number of independent and equally important applications of defeasible logics. The applications fall into three categories: The first has to do with epistemic rules; the second with epistemic requirements; and the third with reasoning policies. Accordingly, the dissertation is divided into three parts too.

Part 1 (Chapters 1–2) is concerned with simple epistemic rules, such as the ones expressed by the statements "If you perceives that X, then you ought to believe that X" and "If you have outstanding testimony that X, then you ought to believe that X." While it is natural to think that these statements are getting at important truths, and that the rules they express should have some role to play in our accounts of epistemic normativity, anyone who thinks that there at least some epistemic rules immediately faces a problem: They need to explain what happens in cases of epistemic conflicts, or situations where such rules come into conflict—such as the one where you perceive a red object and are also told that it is blue. One popular response to this problem suggests that these rules have implicit hedges, or unless-clauses specifying the conditions under which they fail to apply. Another response suggests that these rules are contributory, or that they do not, in fact, specify what beliefs one ought to have, but only what counts in favor or against having them. Chapter 1 uses a defeasible logic framework to state the problem formally and to devise a model for each of these seemingly very different views on rules. The resulting models are simple, but still expressive enough to capture some central philosophical concepts in a mathematically precise way, including those of epistemic reasons, defeaters, and defeat. Chapter 2 establishes a type of equivalence result between these models, suggesting that the views they captures are much closer than standardly thought. It also explores various ramifications of this result for the two views on rules and claims about them that have been made in the philosophical literature. Finally, it formulates a model of what we might call the "mixed view on rules", according to which rules are

contributory, but can also have hedges, and generalizes the equivalence result by relating the mixed view to the hedged-rules view.

Part 2 (Chapter 3) shifts the focus from rules to epistemic requirements, such as the intuitive and widely accepted "If X is supported by your (total) evidence, then you ought to believe that X and "You ought to believe that X if you believe that your evidence supports X." Chapter 3 is naturally seen as doing two things. First off, it uses another defeasible logic to work out a new solution to an important puzzle about epistemic rationality: In case one's (total) evidence can be misleading about what it itself supports—as many epistemologists think—then the above two requirements can come into conflict, suggesting that there are epistemic dilemmas. The defeasible logic-based solution has a number of attractive features when compared to the other solutions from the literature, even though it does comes with an unorthodox perspective on epistemic requirements, a perspective on which they are defeasible. It's also shown—and this is the second major contribution of this part of the dissertation—that defeasible epistemic requirements can be naturally thought of as epistemic ideals, and that the defeasible logic used in solving the puzzle can be naturally seen as the formal backbone of the "conflicting-ideals view" that David Christensen has been advocating for in his recent work. The idea here is, in effect, to understand this view as a move away from the default metaepistemological position according to which epistemic requirements are strict and governed by a strong, but never explicitly stated logic, toward the more unconventional view, according to which requirements are defeasible and governed by a comparatively weak logic. This illuminates the view and helps counter some common objections to it.

Part 3 (Chapters 4–6) applies ideas from logics for defeasible reasoning and formal argumentation theory to the burgeoning debate about the epistemic significance of peer disagreement. The general aim here is to get a better grip on the intuitively appealing "conciliatory views"—which say, roughly, that you're to become less confident in your well-reasoned opinion on some issue if you learn that an epistemic peer holds the opposite opinion—and, in particular, the behavior of these views in scenarios involving higher-order disagreements, such as disagreements over conciliatory views themselves and disagreements over epistemic peerhood. It is shown how the core idea motivating conciliatory views can be naturally expressed in a certain defeasible logic (default logic) with conciliationism emerging as a well-behaved second-order defeasible reasoning policy, saying roughly the following: If your best (first-order or domain-specific) reasoning suggests that X and it's rational for you to think that an epistemic equal disagrees with you regarding X, then you should withhold judgment on whether X under normal circumstances. The logic-based model delivers intuitive results in a plethora of cases discussed in the literature. It also turns out to be very useful in responding to challenges for conciliatory views. Most importantly, it paves the way to solving the well-known and pressing "problem of self-defeat"—which says, roughly, that conciliatory views self-defeat and issue inconsistent directives in scenarios involving disagreements over their own truth. (Also, in the course of pursuing its more philosophical goals, this part of the dissertation devises an intuitive formal argumentation theory framework and shows how it extends a version of default logic. The result is an instance of a known general theorem, but the framework should be of independent interest due to its intuitive character, continuity with a user-friendly version of default logic, as well as other prospective applications.)