

Internal Reasons and the Boy Who Cried Wolf*

Samuel Asarnow

Reasons internalists claim that facts about normative reasons for action are facts about which actions would promote an agent's goals and values (perhaps moderately idealized). Reasons internalism is popular, even though paradigmatic versions have moral consequences many find unwelcome. This article reconstructs an influential but understudied argument for reasons internalism, the "if I were you" argument, which is due to Bernard Williams and Kate Manne. I raise an objection to the argument and argue that replying to it requires reasons internalists to accept controversial metaethical or epistemological commitments with which their theory has not traditionally been associated.

Bernard orders gin but is served gasoline. He takes a sip and spits in disgust. Two different evaluations of Bernard's sipping are natural. In one sense, sipping made sense: given his beliefs and desires, sipping was the thing to do. If we had been in his shoes, we would have done the same. But there is an equally familiar sense in which Bernard's action is criticizable. If we had been there to advise him, we would have told him not to drink.

Robbing a bank, Philippa realizes that success requires shooting the guard. She shoots and succeeds. Here, too, two evaluations are natural. Philippa's shooting makes sense in a way analogous to Bernard's sipping: given her beliefs and goals, it was the thing to do. If we had been in her shoes, we would have done the same. But again there is a different, equally

* This article's journey to publication has been a long one. For help along the way, I thank Alyssa Bernstein, Michael Bratman, John Broome, Jorah Dannenberg, Jamie Dreier, Janet Folina, Geoff Gorham, Donald Hubin, R. J. Leland, Errol Lord, Anna-Sara Malmgren, Diane Michelfelder, Carlos Núñez, Tyler Paytas, Abelard Podgorski, Grant Rozeboom, Debra Satz, Tamar Schapiro, Mike Titelbaum, and Joel Velasco, as well as audiences at Macalester College, Ohio University, Stanford University, and SLACRR, numerous anonymous referees, and several associate editors at *Ethics*.

Ethics 130 (October 2019): 32–58

© 2019 by The University of Chicago. All rights reserved. 0014-1704/2019/13001-0003\$10.00

familiar, sense in which Philippa's action is criticizable. If we had been there to advise her, we would have told her not to shoot. Better to abandon the robbery than to commit murder.

Commonsense evaluations of agents thus come in at least two types, one more subjective and one more objective. We evaluate whether agents' actions make sense, in light of their goals and values (broadly understood), given what they believe the world to be like. And we evaluate whether agents' actions make sense in light of standards deemed important independent of the agent's goals, given how the world really is.

Many normative theories hold that these commonsense evaluations track important kinds of normative facts. Some theories distinguish facts about rationality from facts about what agents ought to do, or have normative reasons to do. Then Bernard and Philippa are said to do what is rational, but not what they ought to do (or have reason to do).¹ Other theories distinguish two senses of ought (or two senses of normative reason), one subjective and the other objective. Then Bernard and Philippa are said to do what they subjectively ought to do, but not what they objectively ought to do.² Either type of theory holds that both commonsense evaluations track normative facts, though they disagree about what kinds of facts those are.

Contrasting with these theories is the influential thesis of reasons internalism.³ Reasons internalists emphasize the importance of evaluating

1. This is the approach of, e.g., Michael E. Bratman, "Intention, Practical Rationality, and Self-Governance," *Ethics* 119 (2009): 411–43; John Broome, *Rationality through Reasoning* (Oxford: Wiley-Blackwell, 2013); T. M. Scanlon, "Structural Irrationality," in *Common Minds*, ed. Geoffrey Brennan et al. (Oxford: Oxford University Press, 2007), 84–103.

2. Compare Derek Parfit, *On What Matters*, vol. 1, ed. Samuel Scheffler (Oxford: Oxford University Press, 2011); Mark Schroeder, "Means-End Coherence, Stringency, and Subjective Reasons," *Philosophical Studies* 143 (February 1, 2008): 223–48; Kurt Sylvan, "What Apparent Reasons Appear to Be," *Philosophical Studies* 172 (2014): 587–606; Daniel Wodak, "Can Objectivists Account for Subjective Reasons?," *Journal of Ethics and Social Philosophy* 12 (2017): 259–79. Compare also Errol Lord, "What You're Rationally Required to Do and What You Ought to Do (Are the Same Thing!)," *Mind* 126 (2017): 1109–54; and Donald C. Hubin, "The Groundless Normativity of Instrumental Rationality," *Journal of Philosophy* 98 (2001): 445–68, 465, who distinguishes judgments of "rational advisability" from "endorsing an action *sans phrase*." As I explain below, I intend talk of "two senses" of ought to be understood in terms of a contextualist semantic theory for "ought," though it is consistent with the idea that "ought" is lexically ambiguous.

3. The arch reasons internalist is Bernard Williams; see Bernard Williams, "Internal and External Reasons," in *Moral Luck* (Cambridge: Cambridge University Press, 1981), 101–13; Williams, "Internal Reasons and the Obscurity of Blame," in *Making Sense of Humanity* (Cambridge: Cambridge University Press, 1995), 35–45; Williams, "Some Further Notes on Internal and External Reasons," in *Varieties of Practical Reasoning*, ed. Elijah Millgram (Cambridge, MA: Harvard University Press, 2001), 91–98; Williams, "Ought and Moral Obligation," in *Moral Luck* (Cambridge: Cambridge University Press, 1981), 114–23; Williams, "Replies," in *World, Mind and Ethics*, ed. J. E. J. Altham and Ross Harrison (Cambridge: Cambridge University

agents in a hybrid way. They claim that facts about normative reasons are facts about which actions promote an agent's goals and values (again, broadly understood), given how the world really is (not given how she believes it to be). This hybrid evaluation renders a different verdict for Philippa than for Bernard: Philippa did something she had a normative reason to do, whereas Bernard did not.⁴ As I will interpret reasons internalism here, it is a claim about the contours of the extension of the objective ought, or objective normative reasons. Williams's view, for example, is that it is false to say that Philippa had no (objective) normative reason to kill the guard, or perhaps even that she ought not kill the guard.⁵

On this interpretation, many versions of reasons internalism have controversial consequences for moral philosophy, consequences many philosophers have found troubling.⁶ They entail, for example, that morality is not universal, in the sense that not every adult human has normative

Press, 1995), 185–224. For other versions of internalism, see Christine M. Korsgaard, "Skepticism about Practical Reason," *Journal of Philosophy* 83 (1986): 5–25; Korsgaard, *The Sources of Normativity*, ed. Onora O'Neill (Cambridge: Cambridge University Press, 1996); Michael Smith, "Internal Reasons," in *Ethics and the A Priori* (Cambridge: Cambridge University Press, 2004), 17–42; Smith, *The Moral Problem* (Oxford: Blackwell, 1994); Julia Markovits, *Moral Reason* (Oxford: Oxford University Press, 2014); Kate Manne, "On Being Social in Metaethics," *Oxford Studies in Metaethics* 8 (2013): 50–73; Manne, "Internalism about Reasons: Sad but True?," *Philosophical Studies* 167 (2014): 89–117; Kieran Setiya, *Reasons without Rationalism* (Princeton, NJ: Princeton University Press, 2007); Alan Goldman, *Reasons from Within* (Oxford: Oxford University Press, 2009). Internalism is sometimes called the "desire-based" theory of reasons or "subjectivism" about reasons, though some authors use those terms to name distinct views. Compare David Sobel, *From Valuing to Value* (Oxford: Oxford University Press, 2017).

4. This holds on paradigmatic versions of internalism, such as those of Williams and Manne. However, some versions of internalism idealize an agent's goals and values, so as to render them more moral, and thus may not entail that Philippa has a reason to shoot. Compare Korsgaard, "Skepticism about Practical Reason"; Korsgaard, *Sources of Normativity*; Smith, "Internal Reasons"; Markovits, *Moral Reason*. My focus throughout will be on the former versions of internalism, which appear to have revisionary consequences for moral philosophy. Similarly, while many paradigmatic noninternalists (that is, externalists) deny that Philippa has a reason to shoot, the claim that she has a reason to shoot is consistent with externalism.

5. Compare Williams, "Ought and Moral Obligation." For some evidence supporting a slightly different interpretation, see Williams, "Internal Reasons and the Obscurity of Blame," 39; Manne, "Internalism about Reasons," 100, 113.

6. Hence Manne's suggestion that reasons internalism is "sad, but true." See Manne, "Internalism about Reasons." Note that, as mentioned above, I am here setting aside the highly ambitious versions of reasons internalism that attempt to avoid these consequences. Compare Korsgaard, "Skepticism about Practical Reason"; and Smith, *Moral Problem*. For discussions of analogous issues in the context of desire-based theories of reasons, see, e.g., Mark Schroeder, *Slaves of the Passions* (Oxford: Oxford University Press, 2007); David Sobel, "Subjectivism and Reasons to Be Moral," in *From Valuing to Value* (Oxford: Oxford University Press, 2016), 16–42. In other work, Manne has also explored a noninternalist theory of reasons that seeks to avoid some of these consequences; see Kate Manne, "Democratizing Humeanism," in *Weighing Reasons*, ed. Errol Lord and Barry Maguire (Oxford: Oxford University Press, 2016), 123–40.

reasons to do what is morally required. They also entail that some people have normative reasons to do seriously immoral things. Strikingly, if reasons internalism is true, these revisionary consequences for commonsense morality follow simply from reflection on the nature of normative reasons.⁷

While controversial, reasons internalism has been influential, as it has been thought to be supported by a number of persuasive arguments. Two such arguments have been widely discussed. In metaethics, it is sometimes noted that some versions of naturalist realism resemble reasons internalism. The familiar attractions of naturalist realism are then used to motivate reasons internalism.⁸ In action theory, reasons internalism is sometimes said to be supported by the idea that normative reasons must be “apt” to serve as motivating reasons (that is, the reasons appealed to in action explanations). Given a broadly causal theory of action, reasons internalism promises to explain how motivating reasons and normative reasons are related.⁹

A third kind of argument for reasons internalism, discussed by Bernard Williams and Kate Manne, has been influential but has rarely garnered critical attention. I call it the “if I were you” argument.¹⁰ It

7. Compare Williams, “Ought and Moral Obligation”; Gilbert Harman, “Moral Relativism Defended,” *Philosophical Review* 84 (1975): 3–22.

8. These attractions include its compatibility with substantive naturalism and its ability to explain the supervenience of the normative on the nonnormative. See, e.g., Schroeder, *Slaves of the Passions*; Goldman, *Reasons from Within*; Smith, *Moral Problem*. Related ideas can be found in Peter Railton, “Moral Realism,” *Philosophical Review* 95 (1986): 163–207; Connie S. Rosati, “Internalism and the Good for a Person,” *Ethics* 106 (1996): 297–326.

9. See, e.g., Williams, “Internal and External Reasons”; Korsgaard, “Skepticism about Practical Reason”; Korsgaard, *Sources of Normativity*; Smith, *Moral Problem*; Julia Markovits, “Internal Reasons and the Motivating Intuition,” in *New Waves in Metaethics*, ed. Michael Brady (New York: Palgrave Macmillan, 2010), 141–65. I discuss my own response to this argument in Samuel Asarnow, “Rational Internalism,” *Ethics* 123 (2016): 147–78.

10. The key ideas in the argument can be found in Williams, “Internal and External Reasons”; Williams, “Internal Reasons and the Obscurity of Blame.” It is more clearly presented as a distinctive argument for internalism in Manne, “Internalism about Reasons.” Related ideas can be found in Julia Markovits, “Why Be an Internalist about Reasons?,” *Oxford Studies in Metaethics* 6 (2010): 255–79. The argument is mentioned in many places but rarely discussed in depth. See, e.g., Russ Shafer-Landau, *Moral Realism: A Defence* (Oxford: Oxford University Press, 2003), chap. 7; Jonathan Dancy, *Practical Reality* (Oxford: Oxford University Press, 2000), chap. 7; Joshua Gert, *Brute Rationality* (Cambridge: Cambridge University Press, 2004); T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), appendix; Rachel Cohon, “Are External Reasons Impossible?,” *Ethics* 96 (1986): 545–56; Christopher Cowley, “A New Defence of Williams’s Reasons-Internalism,” *Philosophical Investigations* 28 (2005): 346–68; Elijah Millgram, “Williams’ Argument against External Reasons,” *Nous* 30 (1996): 197–220; Byron Williston, “Reasons for Action and the Motivational Gap,” *Journal of Value Inquiry* 39 (2005): 309–24; Ernesto V. Garcia, “Value Realism and the Internalism/Externalism Debate,” *Philosophical Studies* 117 (2004): 231–58; Smith, “Internal Reasons”; Stephen Finlay, “The Obscurity of Internal Reasons,” *Philosopher’s Imprint* 9 (2009): 1–22; Markovits, “Why Be an Internalist about Reasons?”; Julia Driver, *Consequentialism* (New York: Routledge, 2012), chap. 6.

describes a distinctive and important kind of thinking, which Williams calls thinking in the “if I were you” mode and Manne calls “reasoning with” a person. Appeals to facts about normative reasons are said to be especially suited for use in this kind of thinking, and reasons internalism is said to follow from that fact.

Among arguments for reasons internalism, the “if I were you” argument is dialectically distinctive. Rather than presupposing controversial views in metaethics or action theory, such as naturalist realism or the causal theory of action, this argument appeals only to (purportedly) widely shared normative intuitions. As Williams once put it, this argument does not “depend on a distinction between fact and value.”¹¹ It appears to be an argument from uncontroversial ideas about the nature of normative reasons to highly controversial moral philosophical claims.

The purpose of this article is to study the “if I were you” argument for reasons internalism, in the context of the debate about the subjective and objective modes of evaluation described above. I ask, does it provide an argument for reasons internalism (of the kind that has revisionary consequences for commonsense moral thinking) that does not rely on controversial metaethical or action-theoretic commitments?

My take is pessimistic. I argue that a careful reconstruction of the argument shows that there is an ambiguity in how reasons internalists have conceived of “if I were you” thinking. I then present a *prima facie* argument that the ambiguity should be resolved in a way that renders the argument unsound. This argument turns on an analogy between Williams’s famous case of Owen Wingrave and Aesop’s fable of “The Boy Who Cried Wolf.” I then consider three responses to the *prima facie* argument, arguing that all three require adopting implausible or (at best) highly controversial metaethical or epistemological commitments. The upshot is that the “if I were you” argument does not provide a theoretically uncontroversial motivation for the kind of reasons internalism that has the revisionary consequences for commonsense morality that many of us find troubling.

A key theme throughout the discussion is that the “if I were you” argument treats an agent’s normative and evaluative beliefs differently from their descriptive beliefs (i.e., their nonnormative, nonevaluative beliefs). It claims that thinking about what there is normative reason for an agent to do allows us to correct the agent’s false descriptive beliefs but requires us to hold fixed their normative and evaluative beliefs. This idea plays a crucial role in several key reasons internalist doctrines. But I argue that it’s difficult to see a principled motivation for this idea that does

11. Williams, “Internal Reasons and the Obscurity of Blame,” 37. See also Bernard Williams, *Ethics and the Limits of Philosophy* (Oxford: Routledge, 1985), 124–27; Williams, “The Structure of Hare’s Theory,” in *Philosophy as Humanistic Discipline*, ed. A. W. Moore (Princeton, NJ: Princeton University Press, 2006), 76–85.

not rest on implausible or, at best, controversial ideas in metaethics or epistemology (such as old-fashioned metaethical noncognitivism or indexical relativism). So while my argument is targeted at the “if I were you” argument, its lessons may apply more generally to other arguments for reasons internalism that rely on treating an agent’s normative and evaluative beliefs differently from their descriptive beliefs.

The article is structured as follows. Section I describes a very general version of reasons internalism. Section II reconstructs the “if I were you” argument. Section III presents the *prima facie* objection, and Section IV considers three possible replies.

I. REASONS INTERNALISM

Because reasons internalism has been formulated in a number of different ways, I begin by introducing terminology that allows me to state a very general version of the view and situate it within the logical space described above.¹²

In my terminology, reasons internalism is a claim about a certain kind of agential evaluation fact. Agential evaluation facts are facts about whether possible actions by agents promote specified outcomes, given specified bodies of information.¹³ On many views, the commonsense evaluations of Bernard and Philippa are both judgments about agential evaluation facts: they are evaluations of whether Bernard and Philippa’s actions promote certain outcomes, given certain information.

Classes of agential evaluation facts can be grouped into families. Some classes of agential evaluation facts are facts about which actions promote outcomes determined by the agent’s mind, whereas others are facts about which actions promote outcomes that are not determined by the agent’s mind. Call the former internalist facts and the latter externalist

12. For versions of internalism, see, e.g., Williams, “Internal and External Reasons”; Korsgaard, *Sources of Normativity*; Smith, “Internal Reasons”; Schroeder, *Slaves of the Passions*; Manne, “Internalism about Reasons.” The thesis I will describe is also entailed by some noninternalist versions of subjectivism.

13. Formally, a class of agential evaluation facts can be specified by a quadruple consisting of a set of possible actions, a set of outcomes, a body of information, and a promotion function mapping actions and information to outcomes. Several of these elements have analogues in the standard semantics for deontic modal verbs such as “ought.” See, e.g., Angelika Kratzer, “What ‘Must’ and ‘Can’ Must and Can Mean,” *Linguistics and Philosophy* 1 (1977): 337–55; Angelika Kratzer, “The Notional Category of Modality,” in *Words, Worlds, and Contexts*, ed. Hans-Jürgen Eikmeyer and Hannes Rieser (Berlin: de Gruyter, 1981), 38–74. On the philosophical significance of this part of semantics, see, e.g., Aaron Bronfman and J. L. Dowell, “The Language of ‘Ought,’ and Reasons,” in *The Oxford Handbook of Reasons and Normativity*, ed. Daniel Star (Oxford: Oxford University Press, 2018), 85–112; John Broome, “A Linguistic Turn in the Philosophy of Normativity?,” *Analytic Philosophy* 57 (2016): 1–14. Plausibly, declarative sentences involving deontic modals characteristically express putative agential assessment facts. Note that I use the word “information” nonfactively throughout.

facts.¹⁴ Similarly, some classes of agential evaluation facts are facts about which actions promote outcomes in light of a body of information determined by the agent's mind, whereas others are facts about which actions promote outcomes in light of a body of information determined by how the world really is. Call the former subjective and the latter objective.

This terminology can be used to perspicuously state a number of important theses in normative theory. For example, those who judge Bernard and Philippa to have acted rationally typically accept the following:

Subjective Internalism about Rationality: Facts about which actions an agent is rationally permitted to perform are subjective internalist facts.

That is, they are facts about which actions promote a set of outcomes determined by the agent's mind, in light of a body of information determined by the agent's mind. Proponents of this principle may, of course, disagree about which parts of the agent's mind determine the outcome set (is it her desires, preferences, or intentions?) and the body of information (is it her all-out beliefs, or her credences?). They may also disagree about whether the outcomes and the information are determined by her mind as it actually is, or an idealized version of her mind (say, her desires and beliefs rendered consistent and coherent).

Similarly, subjective consequentialists accept the following:

Subjective Externalism about Morality: Facts about which actions an agent is morally required to perform are subjective externalist facts.

That is, they are facts about which actions promote a set of outcomes not determined by the agent's mind, in light of a body of information determined by the agent's mind.

Especially important in this context is a thesis about the so-called deliberative ought, a sense of ought given pride of place in many normative theories. This is the sense of ought about which, for example, moralists and egoists are arguing when they argue about whether all agents ought to care about the interests of others.¹⁵ Some normative theorists accept the following:

Objective Externalism about Ought: Facts about what agents deliberatively ought to do are objective externalist facts.

14. Note that a theory according to which the outcome set is determined by an idealized version of the agent's mind counts as internalist.

15. On one (controversial) view, the deliberative ought is the ought about which judgment internalism is true and the "enkrasia" norm holds. See Ralph Wedgwood, *The Nature of Normativity* (Oxford: Oxford University Press, 2007), chap. 1; Broome, *Rationality through Reasoning*, chap. 16.

That is, they are facts about which actions promote a set of outcomes not determined by the agent's mind, in light of a body of information determined by how the world really is.¹⁶

Reasons internalism can profitably be seen as a claim about agential evaluation facts. Consider, for example, Williams's version:

Williams's Internalism: There is a normative reason for A to ϕ only if A could reach the decision to ϕ via a sound deliberative route, beginning with her subjective motivational set.¹⁷

Here the reference to the "subjective motivational set" indicates that the outcome set is determined by the agent's mind, and the reference to the deliberative route being "sound" indicates that the body of information is not determined by the agent's beliefs.

Reasons internalism can thus be understood as objective internalism about normative reasons (which I will abbreviate to objective internalism):

Objective Internalism: Facts about what there is normative reason for an agent to do are objective internalist facts.

That is, they are facts about which actions promote a set of outcomes determined by A's mind, given a body of information representing how the world really is.

Three clarifications about objective internalism are important. First, internalists typically hold that the part of the agent's mind that determines the outcome set consists of her motivational states, certainly including her desires, intentions, preferences, and plausibly also her cares or other mental states that constitute a kind of valuing.¹⁸ There is room for disagreement about exactly what kinds of mental states should be included. But, crucially, all versions of reasons internalism I will consider give the agent's normative judgments (i.e., beliefs about normative and evaluative properties) the role of determining the outcome set. Correspondingly, the body of information about the world does not contain any normative propositions.¹⁹ In what follows, I will refer to the relevant set of outcome-determining mental states as the agent's "goals and values."

16. For influential arguments against object externalism about ought, see, e.g., Frank Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," *Ethics* 101 (1991): 461–82.

17. Adapted from Williams, "Internal Reasons and the Obscurity of Blame."

18. Compare Harry Frankfurt, *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988); Michael E. Bratman, "Valuing and the Will," in *Structures of Agency* (New York: Oxford University Press, 2007), 47–67. Compare also Elizabeth Anderson, *Value in Ethics and Economics* (Cambridge, MA: Harvard University Press, 1993).

19. Internalists sometimes make an exception for propositions concerning thick concepts, such as "braveness." See Williams, "Internal and External Reasons," 104–5.

Second, objective internalists may disagree about how much idealization to impose on the agent's goals and values when determining the outcome set, though the views I will be discussing do not idealize the agent's goals and values so much as to ensure convergence.²⁰

Finally, reasons internalism is typically paired with an analogous objective internalism about the deliberative ought.²¹

Reasons internalism is standardly contrasted with reasons externalism, which, in this terminology, is objective externalism:

Objective Externalism: Facts about what there is normative reason for an agent to do are objective externalist facts.²²

This framework, however, highlights that reasons internalism also contrasts with several other views. These include subjective externalism and subjective internalism, according to which facts about normative reasons are facts about which actions promote certain outcomes in light of a body of information determined by the agent's mind. Reasons internalism also contrasts with a kind of pluralist or hybrid view according to which the outcome set is determined in part by the agent's goals and values and in part by things other than the agent's mind.²³

Importantly, these theses should not be understood as claims about the semantics of the English count noun "reason" or modal verb "ought."²⁴ It is highly plausible that "reason" and "ought" can be used in different ways (to express different kinds of putative agential evaluation facts) in different contexts and so require a contextualist semantics.²⁵ For reasons internalism to be an interesting thesis, it must be a claim about the extension of

20. So I will be setting aside views such as those of Korsgaard, *Sources of Normativity*; Smith, *Moral Problem*. Smith, e.g., famously holds that the desires of all rational agents converge.

21. See, e.g., Williams, "Ought and Moral Obligation." Note, however, that some reasons internalists are not internalists about all normative facts. Compare Williams, "Internal Reasons and the Obscurity of Blame," 39; and Manne, "Internalism about Reasons," 94, 100, 113. I return to this point in Sec. II.A.

22. Note that "not determined by the agent's mind" must be interpreted so as to accommodate an objective theory according to which, e.g., agents have normative reasons to seek pleasurable experiences.

23. Compare Ruth Chang, "Grounding Practical Normativity: Going Hybrid," *Philosophical Studies* 164 (January 17, 2013): 163–87; and Asarnow, "Rational Internalism." Compare also Hubin, "Groundless Normativity of Instrumental Rationality."

24. Though Williams once suggested this interpretation; see Williams, "Internal and External Reasons," 101.

25. Compare, e.g., Tim Henning, "Normative Reasons Contextualism," *Philosophy and Phenomenological Research* 88 (2014): 593–624; Gunnar Björnsson and Stephen Finlay, "Metaethical Contextualism Defended," *Ethics* 121 (2010): 7–36; Stephen Finlay, *Confusion of Tongues: A Theory of Normative Language* (Oxford: Oxford University Press, 2016), chap. 4; Matthew Chrisman, *The Meaning of "Ought": Beyond Descriptivism and Expressivism in Metaethics* (Oxford: Oxford University Press, 2015), chap. 2.

the specific sense of “reason” (and “ought”) that externalists and others have found to be especially important in normative theorizing.²⁶ (On many views, this will be the deliberative sense.) If it were to turn out that reasons internalists and reasons externalists were theorizing about different senses of the term “reason,” then the two views would be consistent, and the debate would be uninteresting.

To foreshadow, my argument below will rely on seeing reasons internalism as a part of this more complex logical space.

II. THE “IF I WERE YOU” ARGUMENT

My central focus in this article will be on a line of reasoning in support of objective internalism that has been widely influential, though it has rarely been critically discussed. Its central idea is that facts about normative reasons are the facts that matter (and that it is appropriate to appeal to) in a distinctive kind of thinking, described as thinking in the “if I were you” mode, or “reasoning with” a person. Objective internalism is then said to follow from that claim, once the distinctive kind of thinking is properly understood.²⁷ Discussions of this line of reasoning are often somewhat inchoate, so after explaining the kind of thinking reasons internalists have emphasized, I formulate a (schematic) version of the argument for study.

A. “If I Were You” Thinking

Williams characterized “if I were you” thinking as follows: “One example of [how claims about reasons are used], which is uncontentionally related to questions raised by the internalist view, is given by advice in the ‘if I were you . . .’ mode. Taking other people’s perspective on a situation, we hope to be able to point out that they have reason to do things they did not think they had reason to do, or, perhaps, less reason to do certain things than they thought they had.”²⁸ Williams claims that when I ask what there is a normative reason for you to do, I “take your perspective” or speak in the “if I were you” mode. Since a person’s perspective is not always transparent to her, this kind of advice can be genuinely helpful

26. That is, they must be claims about the extension of “reason” and “ought” sentences made in a specific class of contexts, namely, those that provide an ordering source determined by what Pittard and Worsnip call the “actually true normative standards” and Broome calls the “final ordering”; see John Pittard and Alex Worsnip, “Metanormative Contextualism and Normative Uncertainty,” *Mind* 126 (2017): 155–93, 170; Broome, “Linguistic Turn,” 10.

27. For the argument itself, see Williams, “Internal and External Reasons”; Williams, “Internal Reasons and the Obscurity of Blame”; Manne, “Internalism about Reasons.” That this is a distinctive kind of argument for reasons internalism is clearer in Manne than in Williams.

28. Williams, “Internal Reasons and the Obscurity of Blame,” 36.

and has “*normative force*.”²⁹ As briefly noted above, Williams contrasts “if I were you” thinking with a more impersonal kind of thinking that is associated with an impersonal sense of “ought” and expressions like “it would be better if . . .”³⁰ Such claims, if true, can be consistent with the claim that there is no reason for the person to act in accordance with them.

Manne develops a similar idea in terms of a social practice: “We can naturally think of the reasons for an agent to perform some action as those considerations which would ideally be apt to be cited in favor of that action, when we are *reasoning with* her about what she ought to do.”³¹ For Manne, “reasoning with” someone is a distinctive mode of interpersonal interaction. Reasoning with someone involves taking toward that person what Strawson called the “interpersonal stance” rather than the “objective stance,” engaging with her rather than treating her as an object to be “managed.”³² As I will interpret Manne, reasoning with someone consists in engaging with that person while remaining in the “if I were you” mode.³³

Williams and Manne highlight three features of this kind of thinking. First, it is constrained by the agent’s perspective on the world, in a way that limits the options that one might appropriately suggest to a person, when thinking in this way. Williams and Manne understand an agent’s perspective in terms of her motivational mental states: her desires, intentions, normative judgments, cares, and so on. They understand accessibility from an agent’s perspective in terms of rationality and reasoning: the possibilities it is appropriate to suggest in “if I were you” thinking are those that the agent could rationally choose, or could come to decide on via reasoning, given her current mental states.³⁴ Ruled out are those possibilities the agent could reach only after a nonrational change in her mental states, as with (perhaps) a nonrational conversion experience.³⁵

Second, there may be normative facts that are off-limits in “if I were you” reasoning. Williams and Manne both claim that it can be true that it would be good or bad (in an impersonal sense) for someone to do something, though reference to that fact (or a suggestion that the agent

29. *Ibid.*, 36.

30. *Ibid.*, 39.

31. Manne, “Internalism about Reasons,” 97 (*italics added*).

32. *Ibid.*, 95.

33. Manne compares her view to Williams’s at *ibid.*, 99.

34. Of course, if you (falsely but reasonably) believe that a possibility makes sense from someone’s perspective, it would be excusable (though inappropriate) to suggest it in “if I were you” thinking.

35. On conversion, see John McDowell, “Might There Be External Reasons?,” in *World, Mind, and Ethics*, ed. J. E. J. Altham and Ross Harrison (Cambridge: Cambridge University Press, 1995), 68–85, 74. Compare Manne’s distinction between conversion and “inspiration” at Manne, “Internalism about Reasons,” 107.

perform the relevant action) is inappropriate in “if I were you” thinking.³⁶ Notoriously, Williams claims that “it would be better” if an abusive husband would stop abusing his wife, and yet there might be no reason for him to do so. Suggesting that he stop would thus be inappropriate (strictly speaking) in “if I were you” thinking, though a “proleptic” suggestion along those lines might be excusable.³⁷

Finally, both Williams and Manne provide vivid illustrations of what it is like to make an inappropriate suggestion in “if I were you” thinking. In Williams’s example, Owen Wingrave is a (fictional) pacifist whose family demands that he honor family tradition by joining the military.³⁸ Wingrave is unmoved, owing to his pacifism and his lack of interest in family tradition. Internalists describe Wingrave’s family’s insistence that he join the military as “browbeating” him or (at best) “bluffing.”³⁹ According to Manne, his family members have retreated from the “interpersonal mode” of interaction to the “objective mode” and are “managing” Wingrave rather than “reasoning with” him.⁴⁰ They are not engaging in “if I were you” thinking.

B. *The “If I Were You” Argument*

While Manne admirably makes explicit the distinctiveness of the “if I were you” argument as an argument for reasons internalism, the argument has not (as far as I know) been laid out in a rigorous way. It is thus worth noting that the argument itself proceeds in five distinct steps.

The first step links existence claims about normative reasons with “if I were you” thinking:

- (1) If there is a normative reason for A to ϕ , then the suggestion that A ϕ is appropriate in “if I were you” thinking about A.

Plausibly, internalists also accept the stronger claim that if R is a normative reason for A to ϕ , then R is an appropriate consideration to bring up in “if I were you” reasoning with A. But only (1) is required for this argument.

The next step links “if I were you” thinking with the idea of an agent’s perspective on the world:

- (2) If the suggestion that A ϕ is appropriate in “if I were you” thinking about A, then the decision to do A is accessible from within A’s perspective.

36. See Manne, “Internalism about Reasons,” 100, 113.

37. See Williams, “Internal Reasons and the Obscurity of Blame,” 39.

38. See, e.g., Williams, “Internal and External Reasons,” 106. The example is drawn from Benjamin Britten’s operatic adaptation of a story by Henry James.

39. The latter characterization is Williams’s. I believe that this use of “browbeating” derives from Millgram, “Williams’ Argument against External Reasons,” 205.

40. Manne, “Internalism about Reasons,” 110.

This is the force of the idea of (say) my reasoning about what to do if I were you. It is your perspective that is at issue. Joining the military, for example, is not accessible from Wingrave's perspective.

Next, the idea of accessibility from an agent's perspective is clarified:

- (3) What counts as accessible from A's perspective depends in part on A's goals and values.

In the context of internalists such as Williams and Manne, this represents the idea that it is the agent's goals and values (not the goals and values of the person making judgments about reasons) that matter. As above, an agent's goals and values consist of a set of her mental states, including at least her desires, intentions, and normative judgments, perhaps moderately idealized. Internalists may disagree with each other about exactly which mental states matter and about how much idealization to impose.

This idea is then further clarified:

- (4) If the decision to ϕ is accessible from A's perspective, then ϕ -ing promotes an outcome somehow determined by A's goals and values.

This follows from the way reasons internalists conceive of accessibility from an agent's perspective. As above, "promote" can be understood in different ways by different internalists.

A connection with normative reasons is then inferred:

- (5) So, if there is a normative reason for A to ϕ , then ϕ -ing promotes an outcome determined by A's goals and values.

Finally, a connection is made with the terminology of Section I:

- (C) So, facts about normative reasons are internalist facts: they are facts about which actions promote outcomes determined by A's mind.

This, I take it, is the "if I were you" argument.⁴¹

Strikingly, this is not an argument for objective internalism in particular. It is an argument for some form of internalism, but it does not provide support for objective rather than subjective forms of internalism. The framework of Section I reveals that premises (3)–(5) are incomplete: they do not specify whether suggestions made within "if I were you" thinking

41. Note that the move from (5) to (C) is an inference to the best explanation: the best explanation of why (5) obtains is that the fact identity claim obtains. While one might object to this inference, I will grant it *ad arguendo* here. I thank an anonymous referee for helping me see this point. As I will explain shortly, this is an argument schema rather than an argument.

are relative to what is accessible to an agent given her goals and values and her beliefs, or given her goals and values, given the facts about the world.

As far as I know, this has never been explicitly noted by reasons internalists, likely because their goal has been to refute objective externalism, which this argument (if sound) accomplishes. Importantly, however, thinking about which actions would promote someone's goals and values, given their beliefs, is intelligible and familiar: it is precisely the kind of thinking we engage in when we think about what it would be rational for someone to do, or what they subjectively ought to do. And this is a kind of thinking that many normative theories make conceptual space for, by holding that it tracks facts about rationality or facts about the subjective ought.⁴²

In the next section, I argue that this incompleteness is the root of a problem for the argument. I argue that, given what Williams and Manne have said about "if I were you" thinking, it appears that premises (3)–(5) should be precisified in a way that generates an argument for subjective internalism, not objective internalism. That is, it is an argument for identifying facts about normative reasons with facts about rationality. I argue that this is a serious problem for the "if I were you" argument.

To be clear, neither Williams nor Manne leaves any doubt that they understand "if I were you" reasoning as being relative to information about how the world really is, not to the agent's beliefs. But I will argue that there is a *prima facie* challenge to that idea and (in Sec. IV) that giving a principled motivation for that idea is not easy.

III. THE BOY WHO CRIED WOLF

A. *The Prima Facie Objection*

How should we understand "if I were you" thinking? Internalists highlight that "if I were you" thinking is constrained by an agent's goals and values (perhaps moderately idealized). Is it also constrained by an agent's descriptive beliefs (that is, their beliefs about nonnormative and nonevaluative matters)? Or does it let us abstract away from the agent's descriptive beliefs?

In this section, I consider what Williams and Manne have told us about "if I were you" thinking. I argue that, given their intuitive characterization of "if I were you" thinking, it appears to be constrained by an agent's descriptive beliefs. That is, given their intuitive characterization, they appear to be describing the kind of thinking typically associated with judgments about rationality or the subjective ought. This produces an objection to the argument.

42. This point is in the spirit of Hubin, "Groundless Normativity of Instrumental Rationality."

In particular, I argue that their characterization of “if I were you” thinking supports replacing (3) with the following:

- (3)* What counts as accessible from A’s perspective depends in part on A’s goals and values and in part on A’s descriptive beliefs.

Given (3)*, (4) and (5) are replaced by the following:

- (4)* If the decision to ϕ is accessible from A’s perspective, then ϕ -ing promotes an outcome somehow determined by A’s goals and values, given a body of information determined by A’s descriptive beliefs.
- (5)* So, if there is a normative reason for A to ϕ , then ϕ -ing promotes an outcome determined by A’s goals and values, given a body of information determined by her descriptive beliefs.

We then can infer the following:

- (C)* So, facts about normative reasons are subjective internalist facts: they are facts about which actions promote outcomes determined by A’s mind, given a body of information determined by her mind.

To argue for (3)*, I consider a case involving an agent with imperfect information. I argue that, given what Williams and Manne have told us about “if I were you” thinking, this example seems closely analogous to the case of Owen Wingrave.

Consider Aesop’s fable of the Boy Who Cried Wolf: “A Shepherd’s Boy had gotten a roguy Trick of crying ‘a Wolfe, a Wolfe’ when there was no such Matter, and fooling the Country People with false Alarms. He had been at this Sport so many times in Jest, that they would not believe him at last he was in Earnest: And so the Wolves brake in upon the Flock, and worry’d the Sheep at Pleasure.”⁴³ Suppose the Country People know that wolves are very rare in this season and that the Shepherd’s Boy has been playing tricks lately. They would then be rational to be highly confident that there is no wolf, or even to all-out believe that there is no wolf. Moreover, if checking on the flock is very costly, it might be highly irrational for them to decide to check on the flock, as the expected value of doing so is much lower than that of alternatives.

Does “if I were you” thinking allow us to suggest that the Country People check on the flock? That is, is “if I were you” thinking constrained by what they actually believe? Or does it allow us to suggest possibilities that only make sense given facts the agent in question is rational to disbelieve?

43. Aesop, “Fable #75,” in *Aesop’s Fables*, ed. Sir Roger L’Estrange (1692), <http://mythfolklore.net/aesopica/lestrange/75.htm>.

Reasons internalists have not explicitly considered cases of limited or misleading information. But if we draw on what they have said when characterizing “if I were you” reasoning, it appears that suggesting to the Country People that they check on the flock shares all three of the important features Williams and Manne emphasize in their discussion of Owen Wingrave and related cases.

First, there is a clear sense in which this possibility is not one that the Country People could decide on rationally. Given their goals and values and their beliefs, the rational thing for them to do is decide not to check the flock. Given its very low expected value, checking on the flock would be highly irrational—indeed, for the Country People to decide to do it, they would have to undergo some kind of a nonrational change. They could come to believe, against the evidence, that a wolf’s presence was likely. They could nonrationally change the values they assign to their options, raising the expected value of checking on the flock. Or they could decide to do something that seemed highly irrational, by their own lights.

Second, it feels natural to apply Williams’s and Manne’s distinction between “if I were you” thinking about the Country People and thinking about impersonal goodness and badness. While the Country People cannot rationally decide to check on the flock, surely it would be better if they did so. As in the cases Williams and Manne discuss, it seems intuitive to say that it would be better if the Country People did something that does not make sense from their perspective.

Finally, if we were to insist that the Country People assist the Shepherd’s Boy (without providing them new evidence), our insistence would intuitively seem like a kind of “browbeating” or “bluffing.” It would be an attempt to “manage their behavior” rather than to reason with them. Given their descriptive beliefs and the evidence available to them, checking on the flock simply does not make sense. We might “proleptically” suggest that they check on the flock, but this would be an inappropriate (though excusable) suggestion.

That these three features all intuitively seem to apply to the case of the Country People generates the *prima facie* objection to the “if I were you” argument. These three features are all reasons internalists have said by way of characterizing “if I were you” thinking. That all three appear in the case of the Boy Who Cried Wolf suggests that “if I were you” thinking is constrained by an agent’s descriptive beliefs, as well as her goals and values (understood as her desires, intentions, normative judgments, etc.). Since (3) is the “if I were you” argument’s interpretation of this idea, as applied to goals and values, (3)* seems reasonable as an interpretation of this idea as applied to descriptive beliefs. But given (3)*, the argument supports (C)*, and not only (C). So the considerations motivating the “if I were you” argument support subjective internalism, not objective internalism.

One might object at this point that the Country People clearly do have a desire that would be served by checking on the flock. It would be easy for them to see this, and even though (by their lights) the expected value of doing so would be low, the evident relevance of this action to their desires would make the suggestion that they check on the flock clearly appropriate in “if I were you” reasoning.⁴⁴

This reply is not open to reasons internalists who rely on the “if I were you” argument, however. Note that Owen Wingrave has some desires the satisfaction of which would be promoted by joining the military (such as his desire to end his uncomfortable dispute with his family). Yet reasons internalists who rely on this argument insist that there is no normative reason for him to join the military.⁴⁵ Because of that claim, such reasons internalists must allow that it is possible for there to be no normative reason at all for an agent to do something that would promote the satisfaction of one of their desires.⁴⁶ The fact that checking on the flock would (in a probabilistic sense, at least) promote one of the desires of the Country People thus does not show that they have any normative reason to do it at all.

The argument I have discussed in this subsection is only a *prima facie* objection to the “if I were you” argument. It is not decisive because it relies on intuitions about how to apply sketchy, only partially theorized ideas to hypothetical cases. What the objection does do, however, is raise a challenge for reasons internalists: Can they provide a more substantial account of “if I were you” thinking that would explain why it is constrained by an agent’s goals and values but not by her descriptive beliefs? Can they give principled reasons for saying that facts about normative reasons are relative to the specific kind of “if I were you” thinking they have in mind?

B. Objective Internalism versus Subjective Internalism

I consider how reasons internalists can answer that challenge in Section IV below. Before that, it is worth considering whether reasons internalists might simply accept subjective internalism. I believe that two sets of considerations suggest that proponents of reasons internalism would be unlikely to find that response attractive.

First, objective internalism and subjective internalism have dramatically different extensions, and the former is substantially more plausible.

44. I thank an anonymous referee for highlighting this objection.

45. After all, this is why their claim that there is a reason for Wingrave to join the military counts as browbeating or bluffing, since it is literally false. Compare Williams, “Internal and External Reasons,” 106.

46. For further discussion of this point, see Samuel Asarnow, “The Reasoning View and Defeasible Practical Reasoning,” *Philosophy and Phenomenological Research* 95 (2017): 614–36.

The locus classicus for this idea is Williams's original "gin and petrol" case (glossed above).⁴⁷ Williams claimed that it is strongly intuitive to deny that there is a normative reason for the person to drink the gasoline (and, a fortiori, that their belief that the liquid is gin is a normative reason for them to do so). This idea has since served as a fixed point in the debate. It is hard to see how subjective internalism could vindicate that intuitive verdict.⁴⁸ Relatedly, normative theorists have typically taken normative reasons to be things ordinary speakers can refer to in English with expressions such as "the reason for A to ϕ is that . . ." That expression requires a true complement, and it is hard to see how subjective internalists could explain that property.⁴⁹

Second, the move to subjective internalism would give credence to the idea that the debate between reasons internalists and externalists has been misguided, in a way internalists evidently believe it has not been. If reasons internalists defend subjective internalism and externalists defend objective externalism, then it is plausible that they are engaged in a merely verbal dispute—they are talking about different senses of the English word "reason."⁵⁰ One side suggests that we use it to pick out facts about the subjective ought (or rationality), and the other suggests that we use it to pick out facts about the objective ought. It would then no longer be obvious that reasons internalists genuinely disagree, for example, with those who defend the universality of morality, robbing the view of much of its interest.

IV. REPLIES

In this section, I consider three ways internalists might reply to my *prima facie* argument. Each approach seeks to give principled reasons for the idea that facts about normative reasons are the facts that it makes sense to discuss in the kind of "if I were you" thinking that requires us to (as it were) respect the agent's actual goals and values but not her actual beliefs. I argue that each approach requires adopting commitments in metaethics

47. See Williams, "Internal and External Reasons," 101–3, where Williams derides subjective internalism as the "sub-Humean" theory of reasons. Compare also his claim that facts about normative reasons have a distinctive "*normative force*" (Williams, "Internal Reasons and the Obscurity of Blame," 36). See also Markovits, *Moral Reason*, 6–7; David Sobel, "Subjective Accounts of Reasons for Action," *Ethics* 111 (2001): 463–64.

48. Note that even a version of subjective internalism that moderately idealized the agent's beliefs would be subject to counterexamples of this form.

49. Compare James Pryor, "Reasons and That-Clauses," *Philosophical Issues* 17 (2007): 217–44.

50. Compare Hubin, "Groundless Normativity of Instrumental Rationality," 465; a suggestion in his spirit would be that reasons internalists are theorizing about "rational advisability" and not ought, "*sans phrase*."

or epistemology that are either implausible or highly controversial. The upshot is that each approach renders the “if I were you” argument substantially less dialectically effective than it has typically been thought to be.

A. *Error*

The first reply begins with Williams’s suggestion that the motivation for objective internalism comes from every agent’s desire to believe what is true and not believe what is false.⁵¹ This might inspire a two-stage reply. First, one might hold that “if I were you” thinking allows us to correct, in general, any errors contained in the agent’s mental states. Second, one might hold that while an agent’s descriptive beliefs can err by being false, an agent’s goals and values can err only by being jointly incoherent. “If I were you” thinking thus allows us to correct false beliefs by substituting true beliefs but allows us to correct goals and values only by rendering them coherent. The Country People err in believing that the Shepherd’s Boy is lying, but Wingrave does not err in judging that pacifism trumps family tradition, since that is a normative judgment, not a descriptive belief, and it coheres with his other goals and values.

This reply implicates reasons internalists in an inconsistent triad, which can be resolved only by adopting an implausible or unattractive idea. The first element in the triad is (a) the idea that “if I were you” reasoning about a person allows us to correct any substantive errors a person makes. This is the leading idea of this reply. The second element is (b) the idea that an agent’s normative judgments cannot be corrected in “if I were you” thinking (except to render them coherent with the agent’s other goals and values). As the cases of Wingrave and the abusive husband illustrate, this is a central commitment of the kind of reasons internalism that Williams and Manne endorse. It is why Williams troublingly claims that an abusive husband may have no reason to treat his wife better, if he (falsely) believes that his behavior is good.

The third element in the triad is (c) the idea that normative judgments (i.e., beliefs about normative and evaluative properties) can be true or false and that false normative judgments are in error. This is a common-sense idea that is also vindicated by any plausible metaethical view and that Williams and other reasons internalists have explicitly endorsed.⁵² Since the desire for true beliefs surely applies to normative judgments as well

51. See Williams, “Internal Reasons and the Obscurity of Blame,” 37.

52. Compare, e.g., *ibid.* Note that contemporary “quasi-realist” noncognitivists hold that normative judgments count as beliefs and are true or false in precisely the same (albeit deflationary) sense as ordinary descriptive beliefs. Compare, e.g., Allan Gibbard, *Thinking How to Live* (Cambridge, MA: Harvard University Press, 2003); Simon Blackburn, *Essays in Quasi-realism* (Oxford: Oxford University Press, 1993).

as descriptive beliefs, (c) is inconsistent with (a) and (b). If normative judgments can err by being false, and “if I were you” thinking allows us to correct any errors an agent makes, then (contrary to (b)) “if I were you” thinking allows us to correct an agent’s normative judgments.

Importantly, it is hard to see how reasons internalists pursuing this strategy can give up (b) or (c). For reasons internalists, (b) is nonnegotiable. A form of reasons internalism according to which facts about objective values that an agent does not believe in provide her with normative reasons is reasons internalism in name only.

Nor can a reasons internalist give up (c), the idea that normative judgments can be in error by being false. Old-fashioned versions of metaethical noncognitivism have this implication but are deeply implausible and have been explicitly rejected by reasons internalists.⁵³ And reasons internalists cannot reply by holding that what makes an agent’s normative judgments true or false is their own internal coherence. This is so for two reasons. First, Williams and Manne explicitly deny that idea, since they countenance facts about impersonal goodness and badness, which are not made true by facts about any agent’s goals and values. Second, one agent’s normative judgments about what a second agent has normative reasons to do are made true (according to reasons internalism itself) by the second agent’s goals and values, not by the first’s. The idea that all of an agent’s normative judgments are made true or false by her own mind is metaethical indexical relativism, not reasons internalism.⁵⁴

Finally, could a reasons internalist resolve this tension by giving up (a) and positing a different universal desire? Perhaps what all agents desire is not to avoid error, but rather to have beliefs that are useful for prediction.⁵⁵ It might then be held that true descriptive beliefs, but not true normative judgments, are useful for prediction. The latter claim, however, is evidently false. Even if (as some claim) normative facts are explanatorily inert, prediction requires only evidence, not explanation.⁵⁶ And

53. Williams denied being a noncognitivist and denied that his defense of reasons internalism “depend[s] on a distinction between fact and value” (“Internal Reasons and the Obscurity of Blame,” 37). However, at other times he appeared to flirt with old-fashioned noncognitivism, as when he denied that beliefs about thin evaluative concepts are “descriptive” or “world-guided” (ibid., 37–38). Compare also Williams, *Ethics and the Limits of Philosophy*, 124–27; Williams, “Structure of Hare’s Theory.”

54. Compare James Dreier, “Internalism and Speaker Relativism,” *Ethics* 101 (1990): 6–26, 6. While some reasons internalists have called themselves “relativists,” reasons internalism and indexical relativism are inconsistent.

55. I thank an anonymous referee for suggesting this point.

56. On the explanatory impotence of normative facts, see Gilbert Harman, *The Nature of Morality* (New York: Oxford University Press, 1977), chap. 1. For arguments against it, see Peter Railton, “Facts and Values,” *Philosophical Topics* 14 (1986): 5–31; Joshua Cohen, “The Arc of the Moral Universe,” *Philosophy and Public Affairs* 26 (1997): 91–134; Nicholas Stur-

normative facts are often evidence of naturalistic matters, as when the fact that a regime is deeply unjust is evidence that it will soon collapse.⁵⁷

Perhaps reasons internalists can defend this reply by finding some other way to modify (a), (b), or (c) so as to render them coherent. How to do so is, at best, far from straightforward, and any such account must avoid both old-fashioned noncognitivism and indexical relativism.

B. Evidence

A second reply to the prima facie objection might begin with the observation that in the case of the Boy Who Cried Wolf, there is evidence we could present the Country People that would make it rational for them to check on the flock: we could show them the wolf. By contrast, one might think that there is no (non-misleading) evidence Wingrave's family could show him that would make it rational for him to join the military.

Turning this observation into a reply to the prima facie objection requires introducing four claims: (a) "if I were you" thinking allows us to imagine presenting the agent in question with new (non-misleading) evidence (this is not obviously a consequence of how Williams and Manne have conceived of "if I were you" thinking); (b) whenever an agent has a false belief, there exists (in the relevant sense) evidence of the truth; (c) presenting agents with new evidence always makes it rational for them to form a relevant true belief; and (d) presenting agents with new evidence never makes it rational for them to alter their goals or normative judgments (unless it does so by way of altering their descriptive beliefs). Together, these ideas entail that "if I were you" reasoning allows us to correct all of an agent's descriptive beliefs (since we can always imagine presenting them with evidence that will make it rational for them to acquire the true beliefs) but requires us to respect her goals and values (since exposure to evidence will not change them directly).

How plausible are these ideas? Claims (a) and (b) raise difficult questions about the relevant conception of evidence, which is obscure and will differ substantially from conceptions of evidence familiar to epistemologists.⁵⁸ Working out this reply would require developing a conception of evidence according to which relevant evidence exists in the case of the Country People but not Owen Wingrave. It would also need to explain

geon, "Moral Explanations," in *Morality, Reason and Truth*, ed. David Copp and David Zimmerman (Totowa, NJ: Rowman & Allanheld, 1985), 229–55.

57. In a context where one knows that unjust regimes usually collapse. Compare Railton, "Moral Realism"; Cohen, "Arc of the Moral Universe."

58. Many epistemologists hold, e.g., that evidence consists in mental states. Compare Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000), chap. 9; Thomas Kelly, "Evidence," in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Stanford, CA: Stanford University, 2014).

in what sense there “exists” evidence of every truth, including facts which might seem nomologically inaccessible to us.⁵⁹ And, plausibly, exceptions would need to be made if it turns out that there are true propositions for which there (logically) cannot be evidence. Finally, given this conception of evidence, it would need to be explained why the first idea is in fact motivated. Given that some of the relevant evidence might be nomologically inaccessible to an agent as she actually is, why does “if I were you” thinking allow us to imagine presenting the agent with that evidence?

Even if those questions can be answered, however, there are substantial reasons to resist the two final ideas. I consider each in turn.

The ideal evidence responsiveness of descriptive belief.—Begin with (c), the idea that presenting agents with evidence will always make them rationally able to form the relevant true beliefs, regardless of what they now believe. Call that idea ideal evidence responsiveness. This idea is crucial to the success of the reply, for without it reasons internalists cannot vindicate the idea that “if I were you” reasoning is not constrained by an agent’s beliefs. However, I think that reasons internalists should be wary of appealing to ideal evidence responsiveness, for two reasons.

The first is that ideal evidence responsiveness is highly controversial. On many views, theoretical rationality is deeply path dependent, in the sense that what an agent believes now constrains what she can rationally come to believe in the future. And, plausibly, this is inconsistent with ideal evidence responsiveness. Consider, for example, a Bayesian view according to which rationality requires agents to respond to evidence by updating their prior credence functions via conditionalization. Then, any agent who assigns credence 1 to any false proposition is a counterexample to ideal evidence responsiveness, for conditionalization provides no way for such an agent to alter that credence in response to new evidence. This reply thus requires either rejecting Bayesianism or adopting some kind of controversial epistemological thesis, such as an unorthodox version of conditionalization,⁶⁰ or the regularity requirement.⁶¹ Given that reasons internalism has not typically been thought to have controversial epistemological consequences, I suspect that this would make many reasons internalists uncomfortable.

Second, considerations internal to reasons internalism motivate its rejection (and Williams himself rejected it, possibly for these reasons⁶²).

59. Such as, in Van Inwagen’s example, “the last true proposition asserted by Plato”; Peter Van Inwagen, “The Incompatibility of Free Will and Determinism,” *Philosophical Studies* 27 (1975): 185–99, 190.

60. Compare Michael G. Titelbaum, *Quitting Certainties: A Bayesian Framework for Modeling Degrees of Belief* (Oxford: Oxford University Press, 2014).

61. Which (conjoined with the probability axioms) entails that any agent who assigns 1 to a falsehood is irrational.

62. See Williams, *Ethics and the Limits of Philosophy*, 130–40.

These considerations relate to the idea of an agent's epistemic standards.⁶³ When agents respond to evidence, they are guided not only by their prior descriptive beliefs but also by their epistemic standards: the rules they follow about how to evaluate evidence.⁶⁴ Two agents might share the same prior beliefs and evidence but respond to evidence differently because their epistemic standards differ. Plausibly, some of an agent's epistemic standards are realized by normative judgments: judgments about how we ought to weight different theoretical virtues, or how "fast" we ought to perform induction, or how we ought to trade off believing the truth and avoiding error. These are judgments about epistemic values.

The idea that some of an agent's epistemic standards consist in normative judgments makes trouble for ideal evidence responsiveness. Crucial to the spirit of reasons internalism is the idea that there can be rational disagreements about matters of value, which are not resolvable by rational argument. For Williams, the abusive husband may be fully rational, and we may be fully rational in condemning him, and there may be no way for us to resolve our disagreement by rational argument. Ideal evidence responsiveness, however, asserts that such disagreements cannot arise concerning judgments of epistemic value. Either rationality requires all agents to, say, weigh simplicity against parsimony in the same way, or, given enough evidence, all agents will necessarily come to weigh them in exactly the same way.⁶⁵ A defense of the "if I were you" argument along these lines would thus need to explain why there can be rational, irresolvable disagreement in our judgments about what to do but no such disagreement in our judgments about how to believe. Perhaps an argument for that idea can be given, but I suspect that few reasons internalists would find it attractive.

The evidence unresponsiveness of normative judgment.—The fourth idea this reply relies on, (d), is that normative judgments are not themselves directly responsive to evidence. Providing an agent with new evidence makes it rational for her to change her normative judgments only when it makes it rational to change the descriptive beliefs on which her normative judgments were based. I think that reasons internalists should be uncomfortable relying on this idea, for two reasons.

First, it is unclear what would motivate this claim. It is not, for example, an implication of any mainstream metaethical theory. Nonnaturalist realism, naturalist realism, and contemporary versions of noncognitivism

63. Compare Miriam Schoenfield, "Permission to Believe: Why Permissivism Is True and What It Tells Us about Irrelevant Influences on Belief," *Nous* 48 (2014): 193–218.

64. On a Bayesian view, an agent's epistemic standards are represented by her *ur*-prior credence function.

65. The former is one version of the uniqueness thesis; compare Matthew Kopec and Michael G. Titelbaum, "The Uniqueness Thesis," *Philosophy Compass* 11 (2016): 189–200.

are all compatible with its rejection.⁶⁶ Indeed, nonnaturalists and analytical reductionist naturalists will likely have independent grounds for doubting (d). Old-fashioned noncognitivism does imply that normative judgment is evidence unresponsive, but (as discussed above) few reasons internalists will accept it.⁶⁷

Second, this idea cannot be derived from reasons internalism itself. One might have thought it could be so derived, in the following way. Normative judgments would be evidence-unresponsive if all evidence for them were available a priori. Then every agent would already possess all of the normative evidence, and so agents could not be presented with new evidence that would make it rational for them to change their normative judgments. And one might have thought that reasons internalism implies that all such evidence was available a priori: since reasons internalism holds that facts about what an agent has reason to do (and ought to do) depend on her own goals and values, shouldn't the relevant evidence be available a priori for her?

But this argument faces two problems, both familiar from Section IV.A. First, reasons internalists like Williams and Manne hold that there are some normative facts that are not made true by any agent's goals and values, namely, facts about impersonal goodness and badness. They thus would need a separate explanation of why evidence about those facts is available a priori. Second, reasons internalism implies that one agent's normative judgments about what other agents have reason to do are made true by facts about the goals and values of the other agents, not by the first agent's goals and values. Evidence concerning what other agents have reason to do is thus straightforwardly not available a priori. The idea that all of an agent's normative judgments are made true by that agent's goals and values is indexical relativism, not reasons internalism.

Defending reasons internalism via the evidence reply thus requires, at best, taking on substantial commitments in either epistemology or metaethics. Perhaps (c) and (d) can be defended, but this reply's reliance on them would render the "if I were you" argument much more controversial than it initially promises to be.

C. Identity

The final response to the prima facie objection emphasizes a different contrast between an agent's goals and values and her descriptive beliefs. Christine Korsgaard has influentially argued, in an internalist context, that what there is normative reason for you to do is constrained by what she calls your "practical identity": "a description under which you find

66. See, e.g., Railton, "Moral Realism"; Smith, *Moral Problem*; Gibbard, *Thinking How to Live*, pt. 4; David Enoch, *Taking Morality Seriously* (Oxford: Oxford University Press, 2011), chap. 7.

67. Though compare Williams, "Internal Reasons and the Obscurity of Blame," 37–38.

your life to be worth living and your actions to be worth undertaking.”⁶⁸ This idea might be used to generate a response to the “if I were you” argument in the following way. First, one might argue that “if I were you” thinking is thinking from within the standpoint of another agent’s practical identity. Second, one might argue that an agent’s practical identity is wholly determined by her goals and values (perhaps idealized). An agent’s descriptive beliefs play no role in determining her practical identity. This would explain why “if I were you” thinking respects an agent’s goals and values but not her descriptive beliefs, thus motivating (3) without motivating (3)*.

This reply, however, is neither attractive nor plausible. It is unattractive because it seems only minimally informative. While Korsgaard’s idea of a practical identity has been influential,⁶⁹ its contours remain obscure, and it is unclear what exactly this reply adds to the characterization of an agent’s perspective given by the original “if I were you” argument. Externalists would be reasonable to complain that a practical identity seems simply like a perspective (as described above), with the stipulation that it does not include any descriptive beliefs.

More importantly, however, that crucial stipulation is simply not plausible. Intuitively, it appears that many of a person’s descriptive beliefs can play a role in determining their practical identity, as Korsgaard understands it. Descriptive beliefs about personal histories, personal relationships, religious matters, and group membership all contribute to determining the descriptions under which people find their lives to be valuable. Indeed, all of Korsgaard’s own examples of practical identities involve descriptive beliefs.⁷⁰ And, plausibly, descriptive beliefs can contribute to practical identities even when they are false (as with a false belief that a person is a child’s biological parent).

A reasons internalist could respond by arguing that “if I were you” thinking requires us to respect an agent’s goals and values, along with those of her descriptive beliefs that partially constitute her practical identity, but allows us to correct the person’s other descriptive beliefs. But that response would make “if I were you” thinking incoherent. The descriptive beliefs that partially constitute a person’s practical identity will stand in logical and evidential relations with her other beliefs. Correcting some,

68. See Christine M. Korsgaard, *Self-Constitution: Agency, Integrity, Identity* (Oxford: Oxford University Press, 2009); Korsgaard, *Sources of Normativity*, chap. 3. Compare also Williams, “Internal Reasons and the Obscurity of Blame,” 36–37, 40–44; Manne, “Internalism about Reasons,” 107.

69. Compare Agnieszka Jaworska, “Respecting the Margins of Agency: Alzheimer’s Patients and the Capacity to Value,” *Philosophy and Public Affairs* 28 (1999): 105–38.

70. “You are a human being, a woman or a man, an adherent of a certain religion, a member of an ethnic group, a member of a certain profession, someone’s lover or friend, and so on” (Korsgaard, *Sources of Normativity*, 101).

but not all, of them would require “if I were you” thinking to work with a deeply incoherent set of beliefs. This response does not seem appealing.

Because it is difficult to see how one might conceive of an agent’s practical identity in a way that does not include at least some of her descriptive beliefs, this reply to the *prima facie* argument seems to me to be the least promising of the three under consideration. It is not plausible that an agent’s goals and values, but not her descriptive beliefs, constitute her practical identity.

V. CONCLUSION

The “if I were you” argument holds that facts about normative reasons are especially suited for use in a distinctive mode of thinking. That mode of thinking requires us to take account of an agent’s goals and values, suggesting only possibilities that are somehow accessible from the agent’s perspective. This argument (if sound) supports the disjunction of subjective internalism and objective internalism.

I gave a *prima facie* argument that the intuitive characterization reasons internalists have given of “if I were you” thinking in fact supports subjective internalism, not objective internalism. Reflection on cases of imperfect information suggests that “if I were you” thinking about an agent is constrained by an agent’s descriptive beliefs, as well as by her goals and values. This argument was not decisive, but it challenged reasons internalists to provide principled reasons for claiming that normative reasons are relative to the specific kind of “if I were you” thinking they emphasize. I consider three ways to respond to that challenge and argue that each requires appeal to further ideas that are either implausible or highly controversial, and not traditionally associated with reasons internalism.

I think that three lessons should be drawn from this discussion. The first is that the dialectical appeal of the “if I were you” argument is limited. It is not a purely intuitive argument for reasons internalism. Like other familiar arguments, it requires controversial theoretical commitments in epistemology or metaethics.

Second, this discussion supports skepticism about the first premise of the “if I were you” argument, on the part of both internalists and externalists. Insofar as “if I were you” thinking is a familiar and natural kind of thinking, it is just thinking about what it is rational for an agent to do, or what agents have subjective reasons to do.⁷¹ If there is to be a sound argument for reasons internalism, it is more likely to come from metaethics or action theory than from the considerations that motivate the “if I were you” argument.

71. This suggestion is in the spirit of Hubin, “Groundless Normativity of Instrumental Rationality.”

Finally, and more broadly, the argument challenges proponents of reasons internalism to consider more closely how their theories treat an agent's normative and evaluative beliefs. Unless they are prepared to adopt the view (antithetical to reasons internalism) that "sound deliberation" corrects normative and evaluative errors, they need to provide an account of why normative and evaluative beliefs are different from descriptive beliefs in this context. To be plausible, that account must steer clear of both old-fashioned noncognitivism and indexical relativism. It is not easy to see how that can be done.