https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

SUBJECTIVE RIGHTNESS*

By Holly M. Smith

I. Background

In the early part of the twentieth century, writers on moral philosophy—prominently Bertrand Russell, C. D. Broad, H. A. Prichard, and W. David Ross—noted that there are many situations in which an agent's misapprehension of his circumstances results in an evaluator's feeling pulled to evaluate the agent's actions as *both* morally right and morally wrong.[1] Consider, for example, the following case:

> *Twin Towers I:* Following the crash of an airplane into a skyscraper, security guard Tom, believing that the elevators will cease working, tells office workers to evacuate the building via the stairwell rather than the elevators. In this case, using the stairs takes too long and all the office workers are killed when the building collapses, whereas the elevators remain operational long enough for the employees to have used them to evacuate safely.

* I am grateful for discussion on these topics to participants in my graduate seminar during the spring of 2008, and in particular to Preston Greene, who convinced me that principles of objective rightness might include reference to the agent's beliefs. I am also grateful to the other contributors to this volume (especially Mark Timmons) for helpful discussion, as well as to the participants (especially Evan Williams and Ruth Chang) in the Rutgers University Value Theory discussion group, the participants in Elizabeth Harman's 2009 ethics seminar, the participants in the 2009 Felician Ethics Conference (especially Melinda Roberts), the participants in the 2009 Dartmouth workshop on *Making Morality Work* (Julia Driver, Walter Sinnott-Armstrong, Mark Timmons, and Michael Zimmerman), and to Nancy Gamburd, Alvin Goldman, Preston Greene, and Andrew Sepielli for comments on earlier versions of this essay. Ellen Frankel Paul provided welcome encouragement to clarify a number of key points.

[1] Bertrand Russell, "The Elements of Ethics" (originally published in 1910; reprinted from Russell, *Philosophical Essays*), in *Readings in Ethical Theory,* ed. Wilfrid Sellars and John Hospers, 2d ed. (New York: Appleton-Century-Crofts, 1970), 10–15; C. D. Broad, *Ethics,* ed. C. Lewy (Dordrecht: Martinus Nijhoff, 1985), chapter 3 (from lectures given in 1952–53); H. A. Prichard, "Duty and Ignorance of Fact" (1932), in H. A. Prichard, *Moral Obligation and Duty and Interest* (Oxford: Oxford University Press, 1968), 18–39; W. D. Ross, *Foundations of Ethics* (Oxford: Clarendon Press, 1939), chapter 7. G. E. Moore has an early discussion of the "paradox" in question, but eventually concludes that we should say the action with the best consequences is right, although the person who does it, believing that it will have bad consequences, is to blame for his choice. See G. E. Moore, *Ethics* (1912; New York: Oxford University Press, 1965), 80–83. Henry Sidgwick uses the terms "subjective rightness" and "objective rightness," but uses the first term to refer to the agent's belief that an action is right (a status now often labeled "putatively right"), and the second term to refer to the fact that the action is the agent's duty in the actual circumstances. See Henry Sidgwick, *The Methods of Ethics* (1874; Chicago: The University of Chicago Press, 1907), 206–8.

doi:10.1017/S0265052509990161

© 2010 Social Philosophy & Policy Foundation. Printed in the USA.

When we focus on the actual outcome of Tom's action of telling the employees to use the stairwell, we want to say that it is wrong—indeed tragic, since it results in the avoidable deaths of scores of office workers. But if we focus on what Tom reasonably believes about the employees' options at the time he advises them, we want to say that he did the right thing. Consider also the following case:

> *Twin Towers II:* Following the crash of an airplane into a skyscraper, security guard Joan, believing that the elevators will cease working, but nursing a grudge against her ex-husband who works for High Tower Investments, tells the High Tower employees to evacuate the building via the elevators rather than the stairwell. The employees comply, and reach safety, whereas if they had taken the stairwell the building would have collapsed and killed them.

Here, when we focus on the actual outcome of Joan's action, we want to say that it is right—that she saved the employees' lives. But if we focus on her giving them advice that she thought would result in their death or injury, we want to say that what she did was very wrong.

To resolve the apparent paradox of such actions being judged both right and wrong, moral theorists in the first half of the twentieth century argued that we must recognize several different senses of such moral terms as "right," "wrong," "obligatory," and "permissible." In one sense of "morally right," they argued, we mean something like "the morally best action in the actual circumstances"; while in another sense of "right," we mean something like "the action that is morally most appropriate in light of the agent's beliefs about those circumstances, even if the beliefs are mistaken."[2] Granting that there are different senses of these terms dissolves the paradox arising from our judgment that each security guard's action is both morally right and morally wrong, since the act can be right in one sense but wrong in the other. Moreover, for those theorists disconcerted by the idea that an act might be an agent's duty even though the agent (through ignorance or mistakes about factual matters) doesn't know the act to be his duty, the concept of an act that is best relative to the agent's beliefs identifies a type of duty to which the agent would always have epistemic access, and thus could fairly be held to blame for violating.[3]

[2] These locutions are somewhat misleading, since a morally *right* action is not necessarily the unique morally best action available to the agent, but may be one of several equally good options. However, for simplicity of exposition, I shall often use "right" when "ought to be done" or "obligatory" would be more accurate. I shall also frequently use "objective rightness" or "subjective rightness" to stand in for the objective or subjective moral status of an action more generally speaking. Note that here and throughout this essay, I am speaking only of *all-things-considered* moral status, not *prima facie* or *pro tanto* moral status. Many of the same issues arise for these latter concepts, and much of my discussion can be applied to them.

[3] Prichard forcefully articulates this worry in "Duty and Ignorance of Fact."

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

Many writers have come to see the importance of making such a distinction. The terms used for expressing these different senses of "right" and "wrong" have varied, but contemporary usage has coalesced around the term "objectively right" for the first sense, and "subjectively right" for the second. The use of these terms to express this distinction is now widely, although not universally, accepted among moral philosophers.[4]

The gap between what is best in light of the actual circumstances of an action, and what is best in light of the agent's beliefs about the circumstances, arises most visibly in the context of consequentialist theories in which the long-term consequences of an action—not readily knowable by its agent—determine its moral status. But this gap can easily arise for deontological or nonconsequentialist theories as well, since the relevant *circumstances* or *nature* of an action (apart from its consequences) may also be difficult for the agent to ascertain accurately.[5] Deontological theories

[4] See, for example, the following discussions: Tim Mulgan, *The Demands of Consequentialism* (Oxford: Clarendon Press, 2001) (discussing versions of consequentialism): "The objectively right action is always what would have produced the best consequences. . . . The subjectively right action is what seems to the agent to have the greatest expected value" (42); David Sosa, "Consequences of Consequentialism," *Mind,* New Series 102, no. 405 (January 1993): "[T]hey can agree that what he did was, say, 'subjective-right' and 'objective-wrong'" (109); Graham Oddie and Peter Menzies, "An Objectivist's Guide to Subjective Value," *Ethics* 102, no. 3 (April 1992): "The subjectivist claims that the primary notion for moral theory is given by what is best by the agent's lights . . . regardless of what is actually the best. The objectivist claims that the primary notion for moral theory is given by what is best regardless of how things seem to the agent" (512); and James L. Hudson, "Subjectivization in Ethics," *American Philosophical Quarterly* 26, no. 3 (July 1989): "In moral philosophy there is an important distinction between *objective* theories and *subjective* ones. An objective theory lays down conditions for right action which an agent may often be unable to use in determining her own behavior. In contrast, the conditions for right action laid down by a subjective theory guarantee the agent's ability to use them to guide her actions" (221; italics in the original).

Some contemporary theorists use the term "rational" to refer to what I am calling "subjectively right." However, since what it would be rational for an agent to do, or what an agent has reason to do, may be ambiguous in just the same way that what it would be right for an agent to do may be ambiguous, I shall not adopt this terminology. Note, though, that the distinction between subjective and objective rightness arises not just in morality but also in other practical fields, such as law, prudence, etiquette, etc. My discussion will be confined to ethics, but much that is said here can be carried over into these other domains.

Some theorists have introduced the distinction between objective and subjective rightness (or something closely similar), not for the reasons I describe, but to serve other argumentative purposes, such as to address the criticism of utilitarianism that it requires agents to constantly calculate the utilities of their actions and thus diverts them from direct attention to the kinds of pursuits and relationships that make life worthwhile. See, for example, Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Philosophy and Public Affairs* 13 (Spring 1984): 134–171.

[5] Philip Pettit, in his essay "Consequentialism," in Stephen Darwall, ed., *Consequentialism* (Malden, MA: Blackwell Publishing, 2003), points out that many nonconsequentialists assume that the properties of actions they find morally relevant are ones such that the agent will always be able to know whether or not an option will have one of those properties. In Pettit's view, this is not generally so. Hence, according to Pettit, "the non-consequentialist strategy will often be undefined" (ibid., 99). In "Absolutist Moral Theories and Uncertainty," *The Journal of Philosophy* 103, no. 6 (June 2006): 267–83, Frank Jackson and Michael Smith argue that absolutist nonconsequentialist moral theorists cannot define a workable account of what it would be subjectively best to do in light of uncertainty.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

may forbid killing the innocent, lying, committing adultery, convicting innocent defendants, stealing, failing to compensate those whom one has unjustifiably harmed, and so forth. But any given agent may be mistaken as to whether a possible killing victim is innocent, whether the statement he makes is untrue, whether the person with whom he has sexual relations is married, whether the defendant is guilty of the crime, whether an item of property belongs to him or to someone else, or whether a given level of compensation covers the loss. Thus, the pressure to recognize two senses of "right" and "wrong" arises equally for both consequentialist and nonconsequentialist moral theories.

The argument for distinguishing objective from subjective rightness may have originally arisen to deal with the fact that in certain circumstances we are pulled to evaluate an agent's action as paradoxically both right and wrong. However, it quickly became clear that another need is served by this distinction as well. It is commonly held that a—or *the*—main function of moral theories is to guide agents in making decisions about what to do.[6] Suppose the security guards in our cases have moral codes that tell them to save the lives of people in the building.[7] Tom, in *Twin Towers I*, can use his moral code to guide his decision, since, in light of his belief that using the stairwell is the only safe evacuation route, he can infer from his moral code that he ought to direct the employees to use the stairwell. His code can guide his decision even though he is mistaken about the facts, and thus chooses, on the basis of his theory, an act that the theory condemns. But consider the following case:

> *Twin Towers III:* Following the crash of an airplane into a skyscraper, security guard Pete must advise the office workers how best to evac-

[6] For a selection of examples, see Eugene Bales, "Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?" *American Philosophical Quarterly* 7 (July 1971): 256–65; Stephen Darwall, *Impartial Reason* (Ithaca, NY: Cornell University Press, 1983), 30–31; Hudson, "Subjectivization in Ethics"; Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990), 43; Frank Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," *Ethics* 101, no. 3 (April 1991): 461–82; Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), 8; Ron Milo, *Immorality* (Princeton, NJ: Princeton University Press: 1984), 22 ("Our primary purpose in passing judgments on our actions is to enable us to guide our choices about how to act"); Jan Narveson, *Morality and Utility* (Baltimore, MD: The Johns Hopkins Press, 1967), 12; J. J. C. Smart, "An Outline of a System of Utilitarian Ethics," in J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973), 44, 46; Michael Stocker, *Plural and Conflicting Values* (Oxford: Clarendon Press, 1990), 10; Mark Timmons, *Moral Theory: An Introduction* (Lanham, MD: Rowman and Littlefield, 2002), 3; and Bernard Williams, "A Critique of Utilitarianism," in Smart and Williams, *Utilitarianism: For and Against,* 124.

[7] Throughout this essay, I will talk about "theories," "principles," and "codes" of objective and subjective rightness. A particularist would reject such generalized statements of what makes actions right or wrong. Nonetheless, the particularist, too, will have to deal with problems arising from agents' mistakes and uncertainties, so he will need to attend to the issues addressed in this essay—something that appears to have been little discussed among particularists.

> uate. Pete believes there is an 80 percent chance that the elevators will become inoperative before they reach the ground floor, and also believes there is a 50 percent chance that people evacuating via the stairwell will not escape the building before it collapses. Pete directs the employees to use the stairwell, but descending takes too long and all the employees are killed when the building collapses. The elevators, however, remain operational long enough for the employees to have used them to evacuate safely.

In this case, Pete cannot use his moral code to make a decision, because it simply tells him to save the lives of people in the building; it tells him nothing about what to do when the probability of saving their lives is less than 100 percent. However, since advising the employees to use the stairwell has a 50 percent chance of saving their lives, whereas advising them to use the elevators has only a 20 percent chance of saving their lives, there is a clear sense in which Pete's advising them to use the stairwell is the better action. Unfortunately, there is an equally clear sense in which this action is the worse action, since it leads to the employees' death, whereas advising them to use the elevators would have saved their lives. *Twin Towers III* demonstrates that the concept of "subjective rightness" can usefully serve a second function: it can be used to pick out the action that it would be wise to perform even though the agent cannot derive guidance directly from his moral code. For Pete, telling the employees to use the stairs is the subjectively right action, while telling them to use the elevators is subjectively wrong; he can decide what to do by choosing the action that has the superior subjective status. Consideration of cases such as this led theorists to recognize that the original distinction between objective and subjective rightness could be leveraged: the concept of subjective rightness can be utilized to provide the moral guidance needed by agents who are *uncertain* (as opposed to *mistaken*) about the circumstances or consequences of their actions, and who therefore need some standard beyond objective rightness in deciding what to do. The frequent uncertainty that agents have about the objective moral status of their prospective actions means that many agents are unable to use their moral code directly to make decisions about what to do. Ideally, the concept of subjective moral rightness dissolves this problem: for every agent who is capable of making a moral decision, on each occasion for decision-making there will be some act identifiable by the agent as one that is subjectively right for her to perform. If she looks to morality for guidance, she can choose the subjectively right act even when she cannot identify which act is objectively right.[8]

[8] Occasionally people respond to *Twin Towers III* by saying, "Of course Pete can use his moral code to make his decision, since it tells him to save the lives of the people in the building, or to choose the method that has the greatest chance of saving their lives." But

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

Of course, the concepts of moral rightness and wrongness are heavily linked to the concept of moral blameworthiness.[9] Once the distinction between objective and subjective rightness/wrongness is recognized, it becomes natural to say something like, "An action is blameworthy only if it is subjectively wrong." An action can be *objectively wrong* but still not blameworthy, as *Twin Towers I* and *III* show: security guards Tom and Pete are not blameworthy for directing the employees to use the stairwell, even though their actions are objectively wrong. One promising way to explain why these actions are not blameworthy is to invoke the concept of subjective rightness, and say that the actions are subjectively right, and hence not blameworthy, even though they are objectively wrong.

Thus, it appears as though the distinction between objective and subjective rightness is an extremely valuable contribution to moral theory. However, it might be claimed that we do not really need this distinction—that we can do all the work we want to do with a more limited set of moral concepts that includes objective rightness/wrongness and blameworthiness/praiseworthiness, but not subjective rightness/wrongness. Thus, it might be claimed that we can say all we need to about *Twin Towers I* and *II* by saying that Tom's act is objectively wrong but not blameworthy (because he believed his act to be objectively right, and had the excuse of ignorance), while Joan's act is objectively right but still blameworthy (because she believed her act to be objectively wrong, but nonetheless chose it). But we cannot say all we need to about *Twin Towers III* by using just these concepts, since we need some way to articulate the moral appropriateness of Pete's act of advising the employees to use the stairwell rather than the elevators (even though this leads to their deaths). It is true that Pete's act is objectively wrong, and that he is not blameworthy for this act. But we cannot explain why he is not blameworthy by saying that Pete believes his act to be objectively right (as we explain Tom's not being blameworthy). By hypothesis, Pete's moral code does not evaluate his act as objectively right, and Pete himself does not believe that it is objectively right, since he is uncertain which act would satisfy his duty to save lives. Thus, it appears we need the distinction between objective and subjective rightness to articulate the moral status of the choice-worthy action in cases where the agent is uncertain which action would be objectively best. Once we have accepted the need for the distinction in

Pete's moral code says only that he is to actually save their lives; advice about what he should do when it is uncertain which escape route would have the greatest chance of saving their lives is part of the job of principles of subjective rightness, and shows why we need them. We are so used to thinking in this fashion that we often do not notice we have switched from a judgment about objective rightness to a judgment about subjective rightness. But see also the remarks about "Remodeling" theorists in the text below.

[9] They are also linked heavily to the concept of an excuse, and in particular to the fact that we excuse (not justify) people for their acts done in ignorance, but I will not try to spell out the ramifications of this in the present essay.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

these cases, we can accept its usefulness in cases such as *Twin Towers I* and *II* as well.

While some theorists might have hoped we could make do with just the standard concepts of objective rightness/wrongness and blameworthiness/praiseworthiness, other theorists (let us call them "Remodeling" theorists)[10] have tried to simplify our moral toolbox by abandoning the *traditional* concept of objective rightness/wrongness, and elevating the concept of subjective rightness/wrongness to take its place. On this view, an action can have only one type of rightness or wrongness, but this fundamental status is determined, not by the action's actual circumstances and consequences, but rather by the content of the agent's beliefs about its circumstances and consequences. A Remodeling theorist would say that the only "right or wrong" judgment we need to make about Tom's action in *Twin Towers I* is that it is right because it is the action most appropriate to the agent's beliefs about its circumstances and consequences.[11] According to such theorists, there is no need to go beyond this by evaluating the action in light of its actual circumstances. Theorists who take this stance are often moved by what they take to be the chief function of moral theories, namely, to guide agents' decision-making. They argue that because agents are frequently mistaken or uncertain about the circumstances and consequences of their actions, it is better to eliminate any evaluation that rests on facts that are unknown to them, and focus solely on evaluations that rest on the decision-maker's beliefs about his circumstances, beliefs which are more accessible to him. For example, many utilitarians have proposed that act or rule utilitarianism be formulated in terms of the *expected* rather than *actual* consequences of the act or rule. According to such Remodeling theorists, Tom's act is right in this fundamental sense of "right"; it can guide him in making his decision. There is no need to introduce any additional concept of rightness. The concept of blameworthiness is then tied fairly directly to the "fundamental" wrongness of the action.[12]

[10] This is a term I employ in *Making Morality Work* (manuscript in progress), and represents a change from the terminology I employed in "Two-Tier Moral Codes," *Social Philosophy and Policy* 7, no. 1 (1989): 112–32.

[11] Common variants of this view would stipulate that the action must be most appropriate to the beliefs that a reasonable person would have in the agent's circumstances, or some similar constraint.

[12] Both Prichard, "Duty and Ignorance of Fact," and Ross, *Foundations of Ethics,* are Remodeling theorists. Recent discussions and defenses of Remodeling theories include Hudson, "Subjectivization in Ethics," 221–29; William H. Shaw, *Contemporary Ethics: Taking Account of Utilitarianism* (Malden, MA: Blackwell Publishers, 1999): 27–31; Brad Hooker, *Ideal Code, Real World* (Oxford: Clarendon Press, 2000); Michael Zimmerman, "Is Moral Obligation Objective or Subjective?" *Utilitas* 18, no. 4 (December 2006): 329–61; and Jackson, "Decision-Theoretic Consequentialism." In *Living with Uncertainty* (Cambridge: Cambridge University Press, 2008), Michael Zimmerman provides the most developed contemporary version and defense of this type of theory. Fred Feldman argues, in "Actual Utility, the Objection from Impracticality, and the Move to Expected Utility," *Philosophical Studies* 129 (2006): 49–79, that the

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

I believe that such Remodeling theorists are mistaken, and that we need both the concepts of objective and subjective moral status. It is not the purpose of this essay to argue for this view. However, before we can seriously assess the view of these theorists, we need a firm grasp on the concept of subjective rightness/wrongness, so that we can accurately determine whether it is sensible to elevate this concept in the manner that Remodeling theorists recommend. The aim of this essay is to propose a novel definition for the concept of subjective moral status. I shall review the definitions available in the literature, and argue that the general approach embodied in these definitions is wrongly conceived and must be abandoned in favor of a more fruitful strategy. A successful definition will help us understand the distinction between objective and subjective moral status, will create an important foundation for evaluating proposed substantive principles of subjective rightness, and will provide groundwork for assessing the claims of the Remodeling theorists.

One final clarification: agents can be mistaken or uncertain about normative matters, as well as about matters of non-normative fact. The difficulties facing such agents, and what to say about them, are deep problems.[13] However, in this essay I will focus only on the difficulties arising from agents' mistakes and uncertainty about matters of non-normative fact, and, where necessary, I will assume the agent has the requisite beliefs about normative matters.

II. Defining "Subjective Rightness" and "Subjective Wrongness"

Despite the fact that theorists have converged on the terms "objective" and "subjective" rightness/wrongness to draw the distinction I have described, the definitions proposed for the terms "subjective rightness" and "subjective wrongness" have varied significantly. Clearly, we need to establish acceptable definitions for these crucial concepts.

The terms "subjective rightness" and "subjective wrongness" were introduced to fill gaps in the existing common and philosophical vocabulary. Hence, assessing the adequacy of any proposed definition will not be a matter of simply determining how accurately it reflects common usage, but rather determining whether it fills the perceived gaps in the desired ways. Reflection on the discussion so far suggests certain criteria that any acceptable definition must meet. One complexity we must acknowledge

Remodeling version of act-utilitarianism using expected utility cannot achieve all the goals its advocates have hoped for.

[13] For initial investigations of these problems, see Ted Lockhart, *Moral Uncertainty and Its Consequences* (New York: Oxford University Press, 2000); Jacob Ross, "Rejecting Ethical Deflationism," *Ethics* 116 (July 2006): 742–68; and Andrew Sepielli, "What to Do When You Don't Know What to Do," in Russ Shafer-Landau, ed., *Oxford Studies in Metaethics* IV (Oxford: Oxford University Press, 2009).

is that (a) dissolving the paradoxical tension created by evaluating an agent's action as both right and wrong (for example, in the *Twin Towers I* and *II* cases) may have slightly different requirements from (b) providing an uncertain agent with guidance (for example, in the *Twin Towers III* case). In stating the criteria for an acceptable definition of "subjective rightness," I will give pride of place to the need to provide guidance for a decision-making agent. Thus, the first-person perspective—that of the agent deciding what to do—will predominate.

A. Criteria of adequacy

I shall work with the following set of criteria of adequacy for a definition of "subjective rightness." These are somewhat rough, but are usable for our purposes.

> *Criterion 1. Normative Adequacy:* A definition of subjective rightness should enable us to identify principles of subjective rightness that will accurately assess (given appropriate background information) the subjective moral status of actions, where an action's subjective moral status will often contrast, in an acceptable manner, with its objective moral status. The principles of subjective rightness should classify actions as subjectively right that strike us as ones it would be reasonable or wise for the agent to choose, given the agent's (possibly faulty) grasp of the situation.
>
> *Criterion 2. Domain Adequacy:* The definition of subjective rightness should enable us to identify principles of subjective moral status that assign subjective status to every action that has objective moral status.[14]
>
> *Criterion 3. Guidance Adequacy:* The definition of subjective rightness should endorse a system of principles of subjective rightness from which agents can derive moral guidance in every situation in which they find themselves, even though an agent may be uncertain or mistaken about which actions have the features that would make them objectively right in that situation.[15]

[14] See Holly M. Smith, "Making Moral Decisions," *Noûs* 22 (1988): 89–93, for a detailed discussion of the concepts of "theoretical" and "practical" domains of a moral principle. The statement of Criterion 2 is fairly rough. Moreover, given the possibility discussed in the text below that a non-possible action is subjectively right, we want the domain of principles of subjective rightness to extend *beyond* the domain of principles of objective rightness. In addition, Criterion 2 is too strong, since an agent may be totally unaware that a certain action (under any description) is available to him (for example, he may not believe he can touch his nose with the tip of his tongue, never having tried or even thought about trying to do this); thus, that action might have objective moral status without having any subjective moral status.

[15] As I shall understand the concept of "moral guidance," it includes *permissions* for agents to act in certain ways, as well as demands that they act in certain ways. Almost every

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

Criterion 4. Relation to Blameworthiness: The action classifications arising from the definition of subjective rightness should bear appropriate relationships to assessments of whether the agent is blameworthy or praiseworthy for her act.

Criterion 5. Normative Compatibility: The definition of subjective rightness should be compatible with the full range of plausible theories of objective moral rightness, so that it is possible to identify acts that are subjectively right relative to each such theory.[16]

Criterion 6. Explanatory Adequacy: The definition of subjective rightness should provide illumination about why subjectively right acts are reasonable or wise to perform, why agents should guide their conduct by reference to such acts, and why these acts are linked to accountability.

I shall use these criteria as guides in assessing proposed definitions of "subjective rightness." However, the Guidance Adequacy Criterion requires further comment. What is it to use a normative principle—such as a principle of subjective rightness—to guide one's decision? Consider John, who wants to follow the principle "Always stop at red traffic lights, and always proceed at green traffic lights." He believes that he sees a red light, and forms the desire to stop his car. But things don't go according to plan: perhaps he stops, but the light was actually green, and what he saw was a red beer advertisement; or perhaps the traffic light was red, but his brakes fail and the car doesn't stop. In both these cases, there is an obvious sense in which he has *not* regulated his behavior in accordance with his principle—but there is another obvious sense in which his decision clearly *has* been guided by it. Reflecting on this case, we may draw the following distinction: an agent is able to use a principle as an *internal guide* for deciding what to do just in case the agent would directly derive a prescription for action from the principle if he wanted to, while an agent is able to use a principle as an *external guide* for deciding what to do just in case the agent would directly derive a prescription for action from the principle if he wanted to, and the act whose prescription he would derive in fact conforms to the principle.[17]

situation is one in which there are several equally morally good options, even though there may be many morally bad options that must be avoided.

[16] Note that there may be limits on this. Some otherwise plausible theories of objective rightness may not be compatible with *any* theory of subjective rightness. This is arguably a fault of these theories of objective rightness, not a deficiency in the definition of subjective rightness. See Frank Jackson and Michael Smith, "Absolutist Moral Theories and Uncertainty," for an argument that absolutist nonconsequentialist theories suffer this failing.

[17] See Holly M. Smith, "Making Moral Decisions," 91–92, for discussion of this distinction. The definitions given in the text are overly simple; the definition of being able to use a principle as an internal guide is further refined by Definition (8) in Section V of the current essay.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

Clearly, it would be ideal if every normative principle, whether it be a principle of objective or subjective rightness, could be used by each agent as an external guide for decision-making. But we have already seen that principles of objective rightness fall short of this ideal, and we must be prepared to discover that principles of subjective rightness may fall short of it as well. However, it seems realistic to insist that principles of subjective rightness—which, after all, are designed to guide agents in making decisions when they are mistaken or uncertain about what the governing principle of objective rightness requires of them—should at least be capable of being used as *internal* decision guides. An agent who cannot find any way to translate his moral values into his *choice* of what to do is an agent who cannot find a way to govern his decision by the considerations he deems most relevant. His decision does not express his moral values, and so in an important way undermines his autonomy.[18] Thus, we want principles of subjective rightness to be capable of being used as internal guides to action, even if they cannot successfully be used as external guides to action. I shall interpret the Guidance Adequacy Criterion as requiring that principles of subjective rightness jointly be usable as internal decision-guides in every situation in which an agent must make a decision.

There are, of course, possible principles of subjective rightness which are not usable as internal guides by a given agent here and now, precisely when the agent must make her decision—but would be usable if she had more information, or had more time to reflect on her circumstances, or had the mental acuity to notice that her beliefs entail, via some complex chain of reasoning, that a certain act is the one prescribed by the principle. By the same token, however, a principle of *objective rightness* that may not be usable as an internal guide by a given agent here and now, when she must make her decision, would be so usable if only the agent had more information, or more time to reflect, or greater mental acuity. We need principles of subjective rightness precisely because agents must often make decisions despite their lack of information or time or ability to deliberate further. Principles of subjective rightness are needed precisely to assist agents in deciding what to do in these circumstances. Hence, when we ask whether a given principle of subjective rightness satisfies the Guidance Adequacy Criterion, we should understand the question to be whether agents are able to use that principle of subjective rightness *at the time they are making a decision, with just the intellectual and informational resources they have at hand*—not whether they would be able to use it if they had more time or some idealized set of resources. Of course, an agent may be blameworthy for not having better resources—perhaps she should have researched her decision more thoroughly before having to make it.

[18] For further discussion of this claim, see Holly M. Smith, "Making Moral Decisions," section V. Pekka Väyrynen has picked up and pursued this idea in "Ethical Theories and Moral Guidance," *Utilitas* 18, no. 3 (September 2006): 291–309.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

But we and she want to know what is best for her to decide, given her actual information, however culpably impoverished it may be.

B. Approaches to defining "subjective rightness"

In the literature, there have been four prominent approaches to defining "subjective rightness." Although details vary, these four approaches can be stated as follows:

(1) Act A is subjectively right just in case A is the act most likely to be objectively right; and
A is subjectively wrong just in case A is not the act most likely to be objectively right.[19]
(2) Act A is subjectively right just in case A is the act that has the highest expected value; and
A is subjectively wrong just in case there is some alternative to A that has a higher expected value than A.[20]
(3) Act A is subjectively right just in case A would be objectively right if the facts were as the agent believed them to be; and
A is subjectively wrong just in case A would be objectively wrong if the facts were as the agent believed them to be.[21]

[19] See Russell, "The Elements of Ethics," 12 (". . . the [act] which will probably be the most fortunate . . . I shall define . . . as the *wisest* act"); Smart, "An Outline of a System of Utilitarian Ethics," 46–47 (". . . the 'rational' . . . action . . . is, on the evidence available to the agent, *likely* to produce the best results . . ."); C. I. Lewis, *Values and Imperatives* (Stanford, CA: Stanford University Press, 1969), 35–38 (". . . right if it probably would have the best consequences"), as quoted in Marcus C. Singer, "Actual Consequence Utilitarianism," in Philip Pettit, ed., *Consequentialism* (Aldershot, England: Dartmouth Publishing Company Limited, 1993), 299; Ross, *Foundations of Ethics,* 157; John Hospers, *Human Conduct* (New York: Harcourt, Brace, and World, 1961), 217 (". . . our subjective duty, namely the act which, in those circumstances, was the most likely to produce the maximum good").

[20] See Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), 24–25; William H. Shaw, *Contemporary Ethics,* 27–31 (as a theory of objective rightness); and Timmons, *Moral Theory,* 124.

[21] See Richard Brandt, "Towards a Credible Form of Utilitarianism," in Hector-Neri Castaneda and George Nakhnikian, eds., *Morality and the Language of Conduct* (Detroit: Wayne State University, 1965), 112–14; Richard Brandt, *Ethical Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1959), 365 (". . . 'did his duty' in [the subjective] sense means 'did what would have been his duty in the objective sense, if the facts of the particular situation had been as he thought they were, except for corrections he would have made if he had explored the situation as thoroughly as a man of good character would have done in the circumstances'"); Peter Graham, "'Ought' Does *Not* Imply 'Can'," unpublished manuscript, 2007: 3–4, http://people.umass.edu/pgraham/Home.html; Fred Feldman, *Doing the Best We Can* (Dordrecht: D. Reidel, 1986), 46; Broad, *Ethics,* 141 (". . . we must say that he is under a formal obligation to set himself to discharge what he knows *would* be his material obligation *if* the situation were as he mistakenly believes it to be"); Milo, *Immorality,* 18 ("If the agent is mistaken about a matter of fact, and, if, had the facts been as he supposed, his act would be wrong, then, unless there are excusing conditions, his act is blameworthy and immoral"); and Judith Jarvis Thomson, "Imposing Risks," in William Parent, ed., *Rights, Restitution, and Risk* (Cambridge, MA: Harvard University Press, 1986), 179 (". . . presumably 'He (subjectively) ought' means 'If all his beliefs of fact were true, then it would be the case that he

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

(4) Act A is subjectively right just in case A is best in light of the agent's beliefs at the time he performs A; and
A is subjectively wrong just in case A is not the best act in light of the agent's beliefs at the time he performs A.[22]

I have phrased several of these definitions in terms of what the agent actually believes. But one popular family of variant definitions involves defining "subjective rightness" in terms of what the agent ought to have believed, what a reasonable person in the agent's position would have believed, what the agent would have believed if she had exercised due diligence, what she would have been justified in believing, etc.[23] Thus, Definition (1) might alternatively read: "Act A is subjectively right just in case A is the act that a reasonable agent would believe to be objectively right." For brevity, I will discuss these popular "reasonable belief" variants only in footnotes until Section VI. Each of these definitions, except (4), assumes a background understanding of the concept of "objectively" right/wrong. For the purposes of this essay, I will assume the informal characterization given in Section I: namely, that an action is objectively right just in case the action is the best one in the actual circumstances. However, subsequent discussion will shed some light on this characterization.

C. Definition (1)

Definition (1) states that *an act A is subjectively right just in case A is the act most likely to be objectively right; and A is subjectively wrong just in case A is not the act most likely to be objectively right.* This definition, like Definition

(objectively) ought'"; although note that Thomson doubts there is any subjective sense of "ought"). Note that the American Law Institute's Model Penal Code, Section 2.04(2) provides that the defense of ignorance of fact "is not available if the defendant would be guilty of another offense had the situation been as he supposed. . . ." Cited in Douglas Husak and Andrew Von Hirsh, "Culpability and Mistake of Law," in Stephen Shute, John Gardner, and Jeremy Horder, *Action and Value in Criminal Law* (Oxford: Clarendon Press, 1993), 161.

[22] See Gibbard, *Wise Choices, Apt Feelings,* 42 ("Thus an act is . . . wrong in the subjective sense if it is wrong in light of what the agent had good reason to believe"; note that Gibbard uses the "good reason to believe" formulation of this definition); Prichard, "Duty and Ignorance of Fact," 25 (". . . the obligation depends on our being in a certain attitude of mind towards the situation in respect of knowledge, thought, or opinion"); Ross, *Foundations of Ethics,* 146–47 (". . . when we call an act right we sometimes mean that . . . it suits the subjective features [of the situation]. . . . The subjective element consists of the agent's thoughts about the situation"; see also ibid., 150, 161, 164); Graham Oddie and Peter Menzies, "An Objectivist's Guide to Subjective Value," *Ethics* 102 (April 1992): 512–33, at 512 (". . . is the morally right action the one which is best in the light of the agent's beliefs?"); and Jackson and Smith, "Absolutist Moral Theories and Uncertainty," 270 (". . . we are in fact talking about what a subject ought to do given their epistemic situation.").

[23] For example, Gibbard, *Wise Choices, Apt Feelings,* 42; Brandt, *Ethical Theory,* 365; and Hospers, *Human Conduct,* 217.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

(2), but unlike Definitions (3) and (4), contains a *substantive rule* for determining an action's subjective status.[24]

As a number of writers (but not all) have noticed, Definition (1) must be rejected, because it often delivers an unacceptable appraisal of an act as subjectively right. Consider the following case (a variant of one much discussed in the literature):[25]

> *Strong Medicine:* Patient Ron consults his physician, Sue, about a moderately serious ailment. Sue can treat Ron with either of two drugs. She believes that giving him no treatment would render his ailment permanent; that drug X would cure Ron partially; and that there is an 80 percent chance that drug Y will cure Ron completely, but a 20 percent chance that Y will kill him.

Suppose Sue's moral code tells her to maximize the welfare of her patients. Her choice, then, appears to be as follows, if we supply some reasonable figures as estimates of the welfare of the patient. "Situation S" is the situation in which *if* Ron takes drug Y he will be completely cured, while "Situation S*" is the situation in which *if* Ron takes drug Y he will be killed. Of course, the outcome for Ron if he receives no treatment, or if he takes drug X, is the same whether Situation S or Situation S* obtains.

TABLE 1. *Possible outcomes in* Strong Medicine

Action	Situation S (probability = .80)	Situation S* (probability = .20)
No treatment	Ron continues ill (value = –500)	Ron continues ill (value = –500)
Give drug X	Ron partially cured (value = 100)	Ron partially cured (value = 100)
Give drug Y	Ron is cured (value = 1,000)	Ron dies (value = –25,000)

According to Definition (1), the subjectively right act for Sue is to prescribe drug Y, since it is most likely to be objectively right. Prescribing drug Y has a .80 probability of maximizing Ron's welfare (and so being objectively right), since there is a .80 probability that Situation S will obtain and Ron will be cured—the best possible outcome. Prescribing drug X has only a .20 probability of maximizing his welfare (and so being objectively right), since there is a .20 probability that Situation S* will

[24] Some authors offer Definitions (1) and (2) as definitions of the concepts of subjective rightness/wrongness, while other authors seem to assume (without stating them) some more general definitions of these concepts, and offer (1) and (2) as substantive rules for determining which acts are subjectively right or wrong. My discussion will focus on (1) and (2) as proposed definitions.

[25] Zimmerman, "Is Moral Obligation Objective or Subjective?" 334; Zimmerman takes the example from Jackson, "Decision-Theoretic Consequentialism," 462–63.

obtain, in which case prescribing drug Y would kill Ron (whereas prescribing drug X would partially cure him), and giving him no treatment would also have a worse outcome than prescribing drug X. Giving him no treatment has a zero probability of maximizing his welfare (and so being objectively right).

But clearly this is incorrect: Sue should not run a 20 percent risk of killing Ron in order to possibly achieve a full cure in this case; it would be wiser of her to prescribe drug X, which will not achieve a full cure, but runs no risk of killing him. Definition (1) gives Sue the wrong advice about what choice to make because it fails to take into account *how* bad (or good) the possible outcomes of her actions are, apart from the bare comparative fact that one outcome is better or worse than another. Hence, it is insensitive to the fact that when prescribing drug Y to Ron does not produce the best outcome, it produces an outcome far worse than anything that might be produced by any of the other options. Definition (1) fails the Normative Adequacy Criterion.[26]

D. Definition (2)

Definition (2) states that *an act A is subjectively right just in case A is the act that has the highest expected value; and A is subjectively wrong just in case there is some alternative to A that has a higher expected value than A.* Definition (2) is explicitly formulated to overcome the problem just seen for Definition (1), since it is formulated to take into consideration, not just the probabilities of the various outcomes of an agent's actions, but also how good or bad those options are, beyond the bare comparative fact that they are better or worse than the outcomes that would be produced by another of the agent's alternatives. The "expected value" of an act is the sum of the expected values of each of its possible upshots, where the expected value of an upshot is the value of that upshot, weighted by the probability of the upshot's occurring. Thus, the expected values of Sue's acts in the *Strong Medicine* case would be as follows:

[26] Note that it would not help Definition (1) to rephrase it along "Reasonable Belief" lines as "Act A is subjectively right just in case A is the act which it would be reasonable for the agent to believe to be most likely to be objectively right, and A is subjectively wrong just in case A is not the act which it would be reasonable for the agent to believe to be most likely to be objectively right." Adverting to what it is reasonable (etc.) for the agent to believe does not enable Definition (1) to escape the problem just discussed.

As several writers have noted, there are cases in which an act that is certain to be objectively wrong is nonetheless one of those that would be subjectively right: see Donald Regan, *Utilitarianism and Co-operation* (Oxford: Oxford University Press, 1980), 264–65; and Jackson, "Decision-Theoretic Consequentialism," 462–63. We can see such a case if we add drug Z to *Strong Medicine*, and in Situation S*, drug Z would completely cure the patient, but in Situation S, drug Z would kill the patient (the opposite of drug Y in these situations). Then giving drug X is certain to be objectively wrong, because in Situation S, drug Y would be better, whereas in Situation S*, drug Z would be better.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

TABLE 2. *Expected values in* Strong Medicine

Action	Situation S (probability = .80)	Situation S* (probability = .20)	Overall expected value
No treatment	Ron continues ill (value = –500)	Ron continues ill (value = –500)	–500
Give drug X	Ron partially cured (value = 100)	Ron partially cured (value = 100)	100
Give drug Y	Ron is cured (value = 1,000)	Ron dies (value = –25,000)	–4,200

According to Definition (2), Sue's prescribing drug X to Ron would be the subjectively right act, because it has the highest expected value (100). Prescribing drug Y would be subjectively wrong, because its expected value (–4,200) is less than the expected value of prescribing drug X. This recommendation to prescribe drug X has vastly more intuitive appeal than the recommendation derived from Definition (1). Because of the intuitive appeal of such recommendations, as well as other reasons, Definition (2) has a long history of support from moral philosophers and decision theorists.

However, even though Definition (2) offers an account of subjective rightness that accords well with our intuitive understanding of what makes some acts better choices than others when the agent is uncertain about the actual facts of his situation (and thus satisfies the Normative Adequacy Criterion), it fails to satisfy the Guidance Adequacy Criterion. The Guidance Adequacy Criterion requires that a definition of subjective rightness endorse principles of subjective rightness that provide guidance to an agent who cannot decide what to do because he is uncertain about the facts of his situation. The principle of subjective rightness endorsed by Definition (2) is simply: "An act is subjectively right if it would maximize expected value, and subjectively wrong otherwise." To apply this principle in making a decision, an agent such as Sue need not have certainty about her circumstances; she need not, for example, feel certain that drug Y would cure Ron. But she *does* need to have probability estimates—not mere "possibility" judgments—about the relevant circumstances. Sue, for example, must be able to assign probabilities to drug Y's curing Ron and to drug Y's killing Ron. Moreover, she must have beliefs about the expected values of her various alternatives, which would normally require her to have calculated these values.[27] In a simple case such as *Strong Medicine,* many (although not all) agents could do this. But many of the decisions that agents must make would necessitate their assigning values and probabilities to events

[27] Of course, it is possible that some advisor might simply inform Sue what the expected values of her options are, relieving her of the need to make these calculations. Regrettably, such advisors are thin on the ground for agents making complex decisions.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

about which they have very little notion what their likelihood is, and would involve the agents' making enormously complex calculations to arrive at each action's expected value. It is completely implausible that every agent has beliefs about the required value assignments and probability estimates, or has the time or ability to make these calculations, or, more generally, has the belief about some action that it would maximize expected value, before a decision must be made.[28] We must conclude that Definition (2), although it directly endorses what is often the correct principle to use in selecting an action, nonetheless violates the Guidance Adequacy Criterion, because this principle cannot be used as a guide by many agents who lack the necessary beliefs or ability or time to apply it.[29]

[28] For a graphic description of these problems, see Feldman, "Actual Utility," 49–79. Note that these problems arise whether the definition or principle of subjective rightness is phrased in terms of objective probabilities or subjective probabilities. Even if it is always possible for an agent to elicit his own subjective assignments of probability, he may not have time to do this before a decision must be made. (Of course, an agent might believe that some act would maximize expected value without having made any calculations.)

To be sure, decision theorists have proven that any decision-maker whose decisions conform to certain rationality postulates governing his subjective probability assignments and his choices over uncertain prospects will necessarily choose the action that maximizes his own expected value. For a classic presentation, see R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York: John Wiley and Sons, 1957), chapter 2. But these subjective values and probability estimates are latent dispositions to make choices in certain situations; the agent himself cannot know what these values and estimates are without a good deal of work. Prior to doing that work, he does not have the information necessary to consciously apply the principle advising him to maximize expected value. Moreover, there is no guarantee that his subjective values (revealed by an array of choices) are actually identical to the moral value that he consciously seeks to maximize in making the present decision. In any event, we are interested in providing a decision-maker with *normative advice* on how to proceed in choosing his action. To be told that he will, if rational, inevitably select the action that maximizes his expected value provides him with no moral guidance.

[29] I argue elsewhere that Definition (2) also fails as a general definition of subjective rightness because it is incompatible with moral theories having certain structures (see my *Making Morality Work*, manuscript).

Note that it would not help Definition (2) to be restated in the form of a Reasonable Belief definition as "Act A is subjectively right just in case A is the act that it would be reasonable for the agent to believe has the highest expected value, and A is subjectively wrong just in case there is some alternative to A that it would be reasonable for the agent to believe has a higher expected value than A." Here, too, adverting to what it might be reasonable (justified, etc.) for the agent to believe does not enable Definition (2) to escape the problem just discussed. There may indeed be cases in which the agent's evidence is sufficiently comprehensive that it would be possible to say that the agent (based on that evidence) would be justified in believing that a given act would have the highest expected value. However, there will be many other cases in which the agent's evidence (or the evidence available to him) is not sufficiently comprehensive to justify a belief about which act has the highest expected value. Moreover, at the time a decision must be made, the agent may not believe that he is justified in having any belief about which action would maximize expected value, or may not be able to identify which such belief would be justified (even though he may be so justified). For this reason, too, the agent could not use a principle of subjective rightness endorsed by this version of Definition (2) in order to make his decision.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

E. Definition (3)

Definition (3) states that *an act A is subjectively right just in case A would be objectively right if the facts were as the agent believed them to be; and A is subjectively wrong just in case A would be objectively wrong if the facts were as the agent believed them to be.* Definition (3) works well in cases such as *Twin Towers I.* In that case, in light of security guard Tom's beliefs about the elevators and the stairs, he also believes that he will save the lives of people in the building by directing the employees to evacuate by the stairs rather than the elevators. Were the facts as Tom believed them to be, his act of directing the employees to use the stairs would be objectively right, so this act counts as subjectively right according to Definition (3). This prescription satisfies the Normative Adequacy Criterion in this case, since we feel that this act is the wisest act for Tom to perform (despite the fact that it results in avoidable tragedy). It also satisfies the Guidance Adequacy Criterion, since Tom can use it to decide which act to perform.

Unfortunately, Definition (3) does not meet these criteria in every case. Consider how to apply it to *Twin Towers III,* in which security guard Pete's relevant beliefs are *probabilistic* ones: he believes that there is an 80 percent *chance* that the elevators will become inoperative before they reach the ground floor, and he believes there is a 50 percent *chance* that people evacuating via the stairs will not get out of the building before it collapses. He further believes that directing the employees to the stairs has the greatest *chance* of saving the employees' lives. To apply Definition (3) to Pete's decision requires us to determine what act would have been objectively right if the facts were as Pete believed them to be. This was easy enough in *Twin Towers I,* since we only needed to ask which act would have been objectively right if the facts were as Tom believed them to be (i.e., if he were correct in believing that he would save the employees' lives by directing them to use the stairs). In a case such as *Twin Towers III,* however, it is much less easy to see how to apply Definition (3). What would the "facts" be if they were as Pete believed them to be? We might try to identify a probabilistic "objective fact" corresponding to Pete's belief that directing the employees to the stairs offers the greatest chance of saving their lives. On some views about probability, there are no "objective" probabilistic "facts" such as a probabilistic fact that directing the employees to the stairs has the greatest chance of saving their lives.[30] On these views, we are blocked from applying Definition (3) to Pete's decision, since we cannot determine which act would be objectively right if the "facts" were as Pete believed them to be—there are no such "facts." In this circumstance, Definition (3) fails to satisfy the Nor-

[30] That is, there are no probabilistic facts other than ones in which the probabilities are 1 or 0. But Pete's beliefs cannot be translated into facts such as these.

mative Adequacy Criterion, the Domain Adequacy Criterion, and the Guidance Adequacy Criterion, since it cannot identify any act as the subjectively right act.

On other views about probability, there might be a sense of objective probability according to which directing the employees to the stairs offers the greatest objective chance of saving their lives. On these views, we would apply definition (3) by asking what act would be objectively right if directing the employees to the stairs offers the greatest objective chance of saving their lives. But Pete's moral code, like most moral codes, ascribes objective moral status to an action in virtue of its *non-probabilistic* features: his moral code says that an action is objectively right if it *will actually* save the employees' lives. His moral code says nothing about the objective status of an action that has the *greatest objective chance* of saving their lives. In the context of his moral code, this probabilistic characteristic of the action is morally irrelevant. Hence, his moral code does not provide an assessment of the objective moral status of any of Pete's options as he understands them. Once again, we are blocked from applying Definition (3) to Pete's decision, since it cannot evaluate actions as subjectively right or wrong in the context of a theory of objective moral rightness that does not ascribe moral relevance to probabilistic features of those actions.[31] Definition (3) must be rejected as violating the Normative Adequacy Criterion, the Domain Adequacy Criterion, and the Guidance Adequacy Criterion in the many cases in which agents have probabilistic beliefs about their options.[32]

III. "Best in Light of the Agent's Beliefs"

We have now seen that Definitions (1), (2), and (3) fail to satisfy all the criteria we introduced for evaluating proposed definitions of the concept of subjective rightness/wrongness.

This leaves us with Definition (4), which states that *an act A is subjectively right just in case A is best in light of the agent's beliefs at the time he performs A; and A is subjectively wrong just in case A is not the best act in light of the agent's beliefs at the time he performs A.* Although worrisomely vague, Definition (4) looks promising. Since it states that an act is evaluated for

[31] Similar conclusions hold if we interpret "probability" as "epistemic probability." Thus, Pete's belief might be interpreted as "My credence level is .8 that the elevators will become inoperative." But there is no way to get from the truth of this belief to a conclusion about what would be objectively right for Pete to do, given that his objective moral code simply tells him to save the lives of the people in the building.

[32] Note that it would not help Definition (3) to be restated in the form of a Reasonable Belief theory such as "Act A is subjectively right just in case A would be objectively right if the facts had been as the agent had reason to believe them to be; and A is subjectively wrong just in case A would be objectively wrong if the facts had been as the agent had reason to believe them to be." What the agent has reason to believe, in many cases, will be probabilistic (as in Pete's case), and so will run into the same problems as the original Definition (3).

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

subjective rightness in light of the agent's beliefs, it is at least consistent with the appropriate evaluations of the agents' choices in *Twin Towers I, II,* and *III,* and in *Strong Medicine,* and thus seems likely to satisfy the Normative Adequacy Criterion. Moreover, since agents normally have access to the contents of their beliefs, it appears that agents can ascertain which action is subjectively right, and hence can apply this concept, as characterized by Definition (4), in their decision-making—which enables it to meet the Guidance Adequacy Criterion. Since blameworthiness is clearly a function (at least in part) of what the agent believes about the circumstances of her choice, Definition (4) appears likely to satisfy the Relation to Blameworthiness Criterion. And, since it appears compatible with any plausible theory of objective rightness, it appears likely to meet the Normative Compatibility Criterion as well.

Nonetheless, there are difficulties with Definition (4). Let us examine them.

A. The accessibility of beliefs

Definition (4) states that whether or not an action is subjectively right (or wrong) is a function of the agent's beliefs. For example, in *Twin Towers I,* it suggests that even if Tom's action of directing the employees to use the stairs is objectively wrong, nonetheless this act is subjectively right, because Tom believes that the employees' lives would be saved by their taking the stairs rather than the elevators. For a principle of subjective rightness, built on Definition (4), to meet the Guidance Adequacy Criterion—to guide any agent in making a moral decision—it must be the case that agents *always* have access to their own beliefs. Thus, it must be the case that even though security guard Pete in *Twin Towers III* does not know (or believe) which escape route would actually be best, he does have access to his belief that directing the employees to the stairs has the greatest chance of saving their lives. If what it is subjectively right for him to do is a function of this belief, and he is aware that he has this belief, then he can make a choice based on a principle of subjective rightness that tells him what to do in light of his beliefs. But if Pete is *unaware* or *uncertain* what his relevant beliefs are, then he cannot apply any principle of subjective rightness that meets Definition (4) in deciding what to do. If agents can be uncertain about the existence and content of their relevant beliefs, then any principle of subjective rightness meeting Definition (4) would fail to satisfy the Guidance Adequacy Criterion in those cases.

Or suppose it is possible for an agent to feel certain what his relevant beliefs are, but to be mistaken about this. Imagine that Pete feels certain he believes that directing the employees to the elevator has the greatest chance of saving their lives, even though he actually believes just the reverse. In such a case, it appears that Definition (4) implies that Pete's subjectively

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

right act is to direct the employees to use the stairs, but he would *believe* that his subjectively right act is to direct the employees to use the elevators. This possibility would open up a gap between what is actually subjectively right for the agent and what the agent may conclude is subjectively right: a gap somewhat parallel to the original gap between what is objectively right for the agent to do and what is subjectively right for him to do. It would undermine one of the original attractions of the concept of subjective rightness, which is that it could be used to identify a type of duty to which the agent has infallible access in his decision-making, even though he may be mistaken or uncertain about which act is objectively right.[33]

But *can* agents be uncertain or mistaken about the content and existence of their own beliefs in the way I just proposed? Of course, in many cases, people do have accurate access to their own beliefs. Some philosophers have argued that this is always the case. However, most philosophers and psychologists now hold that a person's own beliefs are not necessarily accessible to that person (or at least accessible in the time available for making a quick decision). Agents may be unaware of, mistaken, or uncertain regarding the existence or content of their beliefs, just as they can be unaware of, mistaken, or uncertain about the consequences of their actions. Many of our beliefs are "tacit" or stored at an unconscious level—many people believe, for example, that their house has a roof, but that belief is not one of which they are typically conscious or aware in the course of their day-to-day activities. Sometimes we are simply mistaken: a person may, without reflection, assume she has a certain belief, but under the right revelatory circumstances discover she does not have the belief at all. For example, a churchgoer brought up in a conventional religious family may believe that she believes in God, but be mistaken about this, as she discovers when challenged about the content and foundation for this belief. Some beliefs are, and often remain, unconscious because we are motivated not to acknowledge them. Someone raised in a racist community may believe that he personally no longer harbors racist beliefs, but he may be mistaken about this. Or, alternatively, he may have become convinced (through attendance at too many diversity workshops) that he *does* harbor racist beliefs, when actually he does not. Our beliefs, in other words, are not "luminous." (A belief is luminous just in case it is true that if we have that belief, we believe that we have that belief.) Nor are we

[33] It might be urged at this point that Definition (4) should be interpreted as identifying the subjectively right act in light of *all* the agent's beliefs—both his beliefs about his alternative actions, and his beliefs about his own beliefs. But this inclusive set of beliefs would seem to generate two inconsistent answers to what action is subjectively right for him (one arising from the content of his beliefs about the circumstances, and one arising from the content of his beliefs about his own beliefs), so this strategy seems likely to fail. Noting this, however, does call our attention to the fact that we may need to restrict the scope of the agent's beliefs that affect which actions are subjectively right and wrong for him. And, of course, an agent's beliefs about his beliefs about his beliefs about his actions can also be mistaken or uncertain.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

infallible with respect to our beliefs. (We are infallible with respect to belief B just in case it is true that if we believe we have belief B, then we do have belief B.)[34]

To see the implications of this for Definition (4), consider the following case:

> *Learning Disability I:* Allison has overwhelming evidence, and in her heart of hearts she recognizes, that her daughter has a significant learning disability. However, she cannot bring herself to consciously face this fact. If asked, she would truthfully say that she believes that she does not believe her daughter to have any disability. She also believes that having her daughter tested would subject her daughter to peer teasing and undermine her self-confidence, but would maximize her happiness if the test were positive and resulted in remedial action. When given the option to have her daughter tested for the disability, Allison declines.[35]

Let us say that the governing principle of objective rightness tells Allison that an act is objectively right just in case it will maximize her daughter's lifetime happiness, and the governing principle of subjective rightness tells Allison that an act is subjectively right just in case she believes the chance of the act's maximizing her daughter's lifetime happiness is no lower than the chance of any alternative act's maximizing her daughter's happiness.[36] Let us assume that on this theory Allison's declining to have her daughter tested is objectively wrong and—according to Definition (4) and the governing principle of subjective rightness—is also subjectively wrong, since in her heart of hearts Allison believes that her daughter has a learning disability and that having her tested will maximize her lifetime happiness. However, since Allison does not recognize all her own beliefs, she does not regard declining to have her daughter tested as either objec-

[34] For a recent influential philosophical discussion of this issue, see Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000), chapter 4. Williamson introduced the term "luminous," which he applies to cases in which we *are in a position to know something*. For a seminal discussion of the different types of (possible) "privileged access," see William Alston, "Varieties of Privileged Access," *American Philosophical Quarterly* 8 (1971): 223–41. Although most philosophers (and almost all psychologists) would agree with my statements in the text, there has long been philosophical controversy over this point.

Note that the debate about whether the content of mental states, and in particular beliefs, is "broad" or "narrow" is relevant here as well. If the content of a belief (say, the belief that water quenches thirst) partly depends on matters *external* to the believer (e.g., whether the common liquid substance is H_20 or XYZ), then clearly an agent can be mistaken or uncertain about these external matters, and thus mistaken or uncertain about the content of the beliefs he holds.

[35] This case is based on one described in Ian Deweese-Boyd, "Self-Deception," *Stanford Encyclopedia of Philosophy* (October 17, 2006), section 3.0, http://plato.stanford.edu/entries/self-deception/.

[36] Of course, we have already seen that such a principle is normatively faulty, but for reasons of simplicity I will use it in this example.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

tively or subjectively wrong—instead, she holds that this action is both objectively and subjectively right.

Allison's psychology might be somewhat different, as described in the following version of the case.

> *Learning Disability II:* Allison has substantial evidence, and in her heart of hearts she recognizes, that her daughter has a significant learning disability. However, she cannot bring herself to consciously face this fact, even though from time to time it strikes her that her daughter is not learning as fast as other children. When push comes to shove, Allison is uncertain what degree of belief she has that her daughter has a disability, or what degree of belief she has that her daughter's learning ability is within the normal range. She believes (and knows she believes) that having her daughter tested would subject her daughter to peer teasing and undermine her self-confidence, but would maximize her happiness if the test were positive and resulted in remedial action. When given the option to have her daughter tested for a disability, Allison is uncertain about what to do.

Allison's declining to have her daughter tested would be objectively wrong and—according to Definition (4) and the governing principle of subjective rightness—would also be subjectively wrong, since in her heart of hearts Allison believes that her daughter has a learning disability and that having her tested would maximize her daughter's lifetime happiness. However, if Allison accepts Definition (4) and the governing principle of subjective rightness, she is uncertain about whether declining to have her daughter tested is subjectively right or wrong, since she knows that subjective wrongness depends on what probabilities she ascribes to the various relevant facts, but she is uncertain about what she believes on this score.

Thus, if Definition (4) is correct in stating that the subjective moral status of an act depends on the agent's beliefs, there can be cases in which an agent (unaware of or mistaken about her beliefs) can be mistaken about an action's subjective status; and there can also be cases in which an agent (uncertain about her beliefs) can be uncertain about an action's subjective status. The fact that agents can be unaware, mistaken, or uncertain about their own beliefs means that Definition (4) fails the Guidance Adequacy Criterion: there are cases in which the agent can derive no moral guidance from the principles of subjective rightness that the definition endorses, even though some of the actions available to her are subjectively right.[37]

[37] Note one complication here. I have described this case, and Allison's beliefs and uncertainties, relative to a particular principle of subjective rightness. But there may be additional

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

B. The moral significance of beliefs

The discussion up to this point has assumed that the moral significance of beliefs arises only because an agent may be mistaken or uncertain about the features of his possible actions that are relevant to their objective moral status—what we may call the "objective right-making" or "objective wrong-making" features of his acts. As we have seen, because of agents' frequent errors and uncertainties regarding objective right- and wrong-making features of actions, it is useful to define a secondary type of moral status that an action may have in virtue of an agent's beliefs. This secondary status—subjective rightness or wrongness—can be used to pick out the action that it would be best for the agent to choose in light of what he believes, even when his beliefs about the action's objective right- and wrong-making characteristics are faulty.

What we must realize, however, is that there are moral views according to which an action's *objective* moral status may be partly or wholly a function of the agent's beliefs.[38] In other words, an action's objective right- or wrong-making features may include the agent's beliefs. For example, on many moral views, *lying* is wrong, where "lying" is defined (roughly) as asserting what the agent believes to be a falsehood with the intention of deceiving his audience.[39] To perform an act of lying requires

principles of subjective rightness that ascribe subjective moral status to actions in light of *different* beliefs, and Allison might be certain what her beliefs about those matters are, even though she is not certain about the beliefs relevant to the principle in the text. Thus, she could be certain about what this second principle tells her it would be subjectively right to do even though she is not certain about what the original principle tells her. In such a case, her uncertainty about some of her beliefs does not stand in the way of her assigning subjective rightness to one of her actions, because she has certainty about other relevant beliefs. As I will argue later in the text, and have argued elsewhere (Smith, "Making Moral Decisions," 98–99), each principle of objective rightness needs to be supplemented by a variety of principles of subjective rightness, since agents often need to make a decision even though they may not have all the beliefs required to apply the favored principle of subjective rightness to their circumstances. Thus, an agent would have to be uncertain (or mistaken) about a great many of her beliefs to be in a position in which she could not ascribe any subjective moral status to her potential actions.

It would be possible to define a Reasonable Belief version of Definition (4), along the following lines: "An act A is subjectively right just in case A is best in light of the beliefs it would be reasonable for the agent to have at the time she performs A; and A is subjectively wrong just in case A is not the best act in light of the beliefs it would be reasonable for the agent to have at the time she performs A." However, this version of Definition (4) also violates the Guidance Adequacy Criterion—indeed, more pervasively than does the original Definition (4)—since agents are often unaware, mistaken, or uncertain about which beliefs it would be reasonable for them to have.

Note finally that the problem for Definition (4) discussed in this section also arises for Definition (3), and for Definition (2) when the agent must assess her own probability and value assignments.

[38] This is denied by Jackson and Smith, "Absolutist Moral Theories and Uncertainty," 269.

[39] How precisely to define "to lie" is a complex and controversial issue. For a survey treatment, see James Edwin Mahon, "The Definition of Lying and Deception," *Stanford Encyclopedia of Philosophy* (published February 21, 2008), http://plato.stanford.edu/entries/lying-definition/.

the agent to have two beliefs: the belief that his assertion is false, and the belief that his assertion will deceive his audience.[40] Other types of acts commonly held to be wrong also involve attitudinal states that, on analysis, turn out to involve the agent's beliefs.[41] Examples include *stealing* (taking possession of property one believes to belong to another) and *committing murder* (acting in a way that one believes and intends will result in the death of another person). The canonical statement of the Doctrine of Double Effect specifies (among other things) that it would be wrong for an agent to intend a bad effect that he believes will bring about a disproportionately larger good effect.[42] We regard certain kinds of "*attempts*"—such as attempting to murder someone—as objectively wrong, and in such cases, too, the agent must have certain beliefs about the possible upshot of his bodily motions for his act to count as an attempt. Similarly, we think that an agent's *risking* certain grave harms is an objectively wrongful act in itself, even if the harms fail to materialize.[43] Finally, some moral codes prescribe or proscribe certain purely mental acts or attitudes that include beliefs: for example, the Ten Commandments tell us to *honor our parents* (which includes believing one's parents are worthy of respect), but not to *covet our neighbor's house* or *wife* (which includes believing that the house or wife belongs to one's neighbor); and Christianity tells us to *have faith* (which involves believing in God).[44]

[40] For a recent discussion of the assumption that intending to do A always involves believing that one will do A, and references to the literature, see Kieran Setiya, "Cognitivism about Instrumental Reason," *Ethics* 117, no. 4 (July 2007): 649–73. On some views, intending only requires the weaker belief that doing X is *likely* to result in one's doing A.

[41] Of course, criminal and tort law typically define disallowed conduct as including a belief element (e.g., in the definitions of fraud and murder).

[42] For a recent defense of the "intentional" version of the Doctrine of Double Effect, see Michael S. Moore, "Patrolling the Borders of Consequentialist Justifications," *Law and Philosophy* 27 no. 1 (January 2008): 35–96, as cited in John Oberdiek, "Culpability and the Definition of Deontological Constraints," *Law and Philosophy* 27 (March 2008): 105–22. Of course, the full Doctrine of Double Effect also refers to the side-effects of the agent's action, and to the means to his goal.

[43] In this case, to risk something involves believing there is a chance it will occur.

[44] Of course, the Biblical command to honor one's parents includes a command to *act* toward them in certain ways (such as obeying them), but it also seems to involve a command to hold a certain attitude toward one's parents. My comments focus on this latter aspect of the commandment.

There are major issues, of course, about whether such mental activities are appropriate objects for moral duties, since it is unclear to what extent an individual can perform (or avoid performing) the activity voluntarily. The requirement that any duty be one that the agent has the ability to perform "on command" is a common but controversial one; this is not the occasion to discuss it further. See Robert Adams, "Involuntary Sins," *The Philosophical Review* 94, no. 1 (January 1985): 3–32; Richard Feldman, "The Ethics of Belief," *Philosophy and Phenomenological Research* 60, no. 3 (May 2000): 667–95; and Pamela Hieronymi, "Responsibility for Believing," *Synthese* 161, no. 3 (April 2008): 357–73, for defenses of the idea that there can be duties or responsibilities to have certain mental states. Of course, some purely mental "activities" do seem to be ones over which we have the same kind of control that we do over bodily actions: on command, one can search one's memory, do mental arithmetic, review the considerations that favor a certain course of action, etc. In matters of belief, one's

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

It could be cogently argued (and I am sympathetic with this argument) that in the case of each of these types of performance (except those that involve purely mental activities, such as believing in God), it is only the underlying non-mental activity that is objectively wrong. On this view, when we evaluate as "right" or "wrong" the more complex act (such as lying or stealing)—an act that involves bodily motions, the surrounding circumstances, *and* the agent's beliefs and desires—we are using a kind of time-saving (but misleading) shortcut that merges together considerations of objective moral status, subjective moral status, and blameworthiness. Thus, in the case of lying, it could be argued that what is genuinely *objectively* wrong is making an assertion that misleads the person who hears it; what is *subjectively* wrong is making an assertion in the belief that it is false and will mislead; and what is *blameworthy* is performing an act that one believes to be subjectively wrong. Similar analyses, identifying an objectively wrong bodily movement (and set of circumstances) at the core of each of these acts, could be offered for stealing, committing murder, and harming someone in order to bring about a good effect.

However, successfully carrying out this program of eliminating reference to any mental aspects when defining objectively wrong actions may not be easy. For example, philosophers who have worked on precise definitions of "lying" have concluded that one can only lie to one's *intended* audience, not to an eavesdropper who happens to overhear and be misled by one's statement.[45] But if we agree that lying must involve misleading an intended audience, we have re-imported into the morally relevant definition of the act a reference to the agent's beliefs about his audience that would be difficult to eliminate. It would also be difficult to provide an eliminative account of "attempting" and "risking" harms.[46] And even if the "elimination" program were successful, it would undeniably fly in the face of the stated content of many commonly accepted moral codes, which incorporate, in the list of activities that are objectively wrong, activities whose definitions undeniably refer to the agent's beliefs. And, of course, such a program could not touch activities, such as coveting one's neighbor's wife or believing in God, which are purely mental and involve beliefs. Finally, if we ask a deeper question about what kinds of human conduct are properly subject to moral evaluation, the answer must include human *acts* as contrasted with mere human *behavior* such as sneezing. But acts are human behaviors that the agent intends to perform (or which are generated by more basic acts that the agent intends to

mental inquiry or search may be controlled, but not one's mental response to the result of the inquiry.

[45] See Mahon, "The Definition of Lying and Deception," for discussion. Clearly, this condition would be deemed to be relevant to the lie's *moral* status; eavesdroppers have no right that they not be misled.

[46] Wrongful acts such as *attempting to harm someone* seem to depend on one's beliefs about what one is doing, not (for example) on the objective probability of one's acting in a way that will harm the person. I thank Preston Greene for pointing this out.

perform). To intend to perform an act, in most cases, involves having certain beliefs, such as the belief that one's moving one's finger will pull the trigger and fire the gun. Thus, the performance of even an unintentional and unforeseen act (such as accidentally killing Jones) requires the agent to have certain beliefs (say, the belief that moving his finger will pull the trigger, fire the gun, and kill the deer that the agent mistakenly believes he sees).

It appears, then, that many activities are commonly deemed to be objectively right or wrong in whole or in part because of the agent's beliefs—in short, that the agent's beliefs are, in some cases, objective right- or wrong-making features of an act. Definition (4), however, defines an act A as subjectively right just in case A is best in light of the agent's beliefs at the time he performs A; and it defines A as subjectively wrong just in case A is not the best act in light of the agent's beliefs at the time he performs A. A deontological moral code that prohibits lying (as it is usually understood) implies that an act of lying is wrong *at least partly in light of an agent's beliefs.* According to Definition (4), it appears as though such a moral code could be interpreted as saying that lying is *subjectively* wrong, whereas the aim of the code is to prohibit lying as *objectively* wrong. The point of an objective code prohibiting lying is not to offer an agent guidance about what it is wisest to do when the agent's grasp of his circumstances is faulty or inaccurate; that is the aim of principles of subjective, not objective, rightness. Exactly how Definition (4), would handle "mixed" acts—ones whose right- and wrong-making features include the agent's bodily movements, surrounding circumstances, *and* the agent's beliefs—is somewhat unclear.[47] However, if a moral code prescribes or prohibits certain beliefs in themselves, such as the belief in God, it seems clear that it would be classified by Definition (4) as a code of subjective rightness. And this seems to be a mistake, since the point of the principle prescribing belief in God is to tell an agent what it is simply best for him to believe—not to tell him what it is best for him to believe in light of the fact that he is mistaken or uncertain about what he believes.

What we are seeing here is that an agent's beliefs might be relevant to the *objective* moral status of an activity, not just to the subjective moral status of that activity. Whether or not an agent believes P is, of course, an "objective" fact, just as whether or not her act would cause pain to someone else is an "objective" fact. Clearly, a moral theory can cogently entail that an agent's beliefs affect the objective moral status of her actions in the

[47] Note that subjectively right/wrong acts themselves are typically understood to have "objective" features in addition to what the agent believes of them: they must be acts that are potentially performable by the agent, not just figments of the agent's imagination. There may be temporal factors as well, linking the time of the action and the time of the agent's beliefs. If this is correct, then Definition (5) (discussed below in Section IV) must apply to acts having mixed "objective" and "subjective" features. But for discussion of this assumption, see the fifth point in my discussion of Definition (5) below.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

same way that the consequences of the action affect its objective moral status, or in the same way that the fact that the action would break a promise affects its objective moral status.[48] Whether and how the agent's beliefs affect the objective status of her actions is an entirely separate question from the question of whether and how the agent's beliefs affect what we are calling the "subjective" status of her actions. One can hold that lying is objectively wrong, and that lying necessarily involves making a statement one believes to be false, without addressing our initial question of how to reconcile apparently conflicting evaluations of an agent who acts from false beliefs about the nature of her act, or our initial question of how a moral code can provide guidance to an agent even though she may be mistaken or uncertain about the nature of her action.

Moreover, it is clear that a moral code that deems an agent's beliefs to be relevant to the objective moral status of her actions needs the distinction between objective and subjective rightness, just as does a moral code that deems only non-mental states to be relevant to the objective moral status of an action. Suppose one accepts that an agent's beliefs affect the objective moral status of her activities (one accepts, for example, that lying is wrong, and that in order to lie one must believe one's statement to be false; or one accepts that faith in God is morally required). As we saw in Section III.A, agents may be unaware of, mistaken, or uncertain regarding the existence or content of their beliefs, just as they can be mistaken or uncertain about the consequences of their actions. The churchgoer brought up in a conventional religious family believes that she has faith in God, but she may be mistaken about this. The person raised in the racist community believes that he no longer harbors racist beliefs, but he may be mistaken about this. Allison is mistaken or uncertain about what she believes regarding her daughter's learning abilities. Thus, a moral code that assesses the objective moral status of actions partly or wholly in terms of the agent's beliefs must confront situations in which what the agent actually believes diverges from what she believes (or is certain) her beliefs are. These are the very kinds of situations that the concept of subjective rightness was invented to handle.

What this means is that we cannot tell, merely by noting that an isolated moral principle ascribes moral status to an action in virtue of the agent's beliefs, whether that principle is a principle of objective or subjective rightness. The schema "An act is morally right if it has features F, G, and H," where at least one of these features involves the agent's beliefs, could be either a principle of objective rightness or a principle of subjective rightness. Content alone will not tell us this, because the agent's beliefs might be right-making for a principle of objective rightness, or right-making for a principle of subjective rightness. What we must recognize is that the concept of subjective moral status always implicitly

[48] I am grateful to Preston Greene, who persuaded me of this point.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

imports a paired concept of objective rightness, relative to which it must be understood.[49] A principle of subjective rightness has to be defined in relation to a foundational principle of objective rightness, and the principle of subjective rightness can only be understood and assessed as appropriate relative to the principle of objective rightness.

IV. Proposed Solution to Defining "Subjective Rightness"

We need to find a definition of subjective rightness that satisfies the six criteria set out in Section II.A and avoids the problems we have noted for the preceding four definitions. I believe the best way to do this is to approach the question somewhat differently. Up until now, we have focused on proposed definitions of an *act's* being subjectively right/wrong. What we need to do instead is to focus first on characterizing what makes a normative *principle* a principle of subjective rightness/wrongness, and then use this definition to characterize when an act is subjectively right/wrong. What makes a normative principle a principle of *subjective* rightness is not its content per se, but rather its relation to some governing principle of objective rightness. The principle of subjective rightness lays out the evaluative status of actions, relative to the principle of objective rightness, for agents who are mistaken or uncertain about whether those actions have the right-making features specified by the principle of objective rightness. Our definition must capture this essential fact.

Our definition of a principle's being subjectively right/wrong also needs to accommodate the fact that a given principle of objective rightness may need to be supplemented by *several* substantive principles of subjective rightness, since a principle of subjective rightness that one agent may be able to apply in one set of circumstances may not be usable by other agents (or by the same agent in a different set of circumstances) when those agents have less rich sets of beliefs about their options. For example, an agent who does not have a rich enough set of beliefs to use a principle prescribing the maximization of expected value might still have a sufficiently rich set of beliefs to use a satisficing principle, or the maximin principle. These principles of subjective rightness can be understood as forming a rough hierarchy. If the agent has a set of beliefs that would enable him to use a more highly ranked principle in this hierarchy, then what is subjectively right for him will be the act prescribed by the more highly ranked principle.[50] Clearly, a normative standard is needed for

[49] This point is further enforced by the fact that many Remodeling theorists have advocated, as principles of *objective* rightness, principles with exactly the same content as principles advocated by others as principles of *subjective* rightness (e.g., "One ought to maximize expected utility"). Examination of the right-making feature identified by this principle does not tell us whether it is a principle of objective or subjective rightness.

[50] I have argued for the necessity of a hierarchy of principles of subjective rightness in my essays "Making Moral Decisions," and "Deciding How to Decide: Is There a Regress Prob-

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

determining what makes one principle "higher" than another, but developing such a standard must be the work of another occasion.

Given these ideas, we can characterize a principle of subjective status (rightness or wrongness) as follows:

Definition (5):
If Q is a principle of objective moral status, and Q stipulates that F is a right-making[51] feature of actions and that G is a wrong-making feature of actions, *then*

(1) A normative principle P is a principle of *subjective rightness* relative to principle Q just in case, for any agent S, either of the following is true:

(A) *if* agent S believes (correctly or incorrectly) of some act A that A is possible for him to perform and that A has feature F, *then* principle P prescribes A, relative to principle Q and relative to S's non-normative beliefs about A; or
(B) *if* (i) agent S believes (correctly or incorrectly) of some act A that A may be possible for him to perform, *and if* (ii) S is uncertain whether any act available to him has feature F, and if so, which act does have F, *then* principle P prescribes A relative to principle Q and relative to S's non-normative beliefs about A; and

(2) A normative principle P is a principle of *subjective wrongness* relative to principle Q just in case, for any agent S, either of the following is true:

(A) *if* agent S believes (correctly or incorrectly) of some act A that A is possible for him to perform and that A has feature G, *then* principle P prohibits A, relative to principle Q and relative to S's non-normative beliefs about A; or
(B) *if* (i) agent S believes (correctly or incorrectly) of some act A that A may be possible for him to perform, *and if* (ii) S is uncertain whether any act available to him has feature G, and if so, which act does have G, *then* principle P prohibits

lem?" in Michael Bacharach and Susan Hurley, eds., *Essays in the Foundations of Decision Theory* (Oxford: Basil Blackwell, 1991), 194–219. For decision theorists' discussions of the need for multiple decision-guides, see Clyde C. Coombs, Robyn M. Dawes, and Amos Tversky, *Mathematical Psychology* (Englewood Cliffs, NJ: Prentice-Hall, 1970), chapter 5; and Michael Resnik, *Choices* (Minneapolis: University of Minnesota Press, 1987), 40.

[51] "Right-making" is here construed as "all-things-considered right-making." A parallel version of Definition (5) could be stated for "prima facie right-making" (and similarly for "wrong-making").

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

A relative to principle Q and relative to S's non-normative beliefs about A.[52]

Definition (5) is intended to capture several ideas: (i) that we need to focus first on a definition of what makes a moral principle a *principle* of subjective moral status (rather than objective status) before moving on to say what makes a given *act* subjectively right; (ii) that a principle of subjective status has that standing *relative to* some principle of objective status, not simply taken in isolation; (iii) that when an agent believes that some act has a right-making (or wrong-making) feature identified by the principle of objective status, then an agent can use the principle of objective status internally to make a decision, so that it can serve as a principle of subjective status; and (iv) that provision must be made for the fact that in cases of uncertainty, there may be more than one principle of subjective status available to decision-making agents.

Thus, for example, suppose Q is a principle of objective status stating that an action is wrong if it involves killing an innocent person. One possible subordinate principle P states that when an agent is uncertain about whether act A would involve killing an innocent person, it would be wrong (relative to Q, and to the agent's non-normative beliefs) for her to perform act A if she believes that the action has a probability greater than .001 of killing an innocent person. Principle P qualifies as a principle of subjective wrongness relative to Q. To say that principle P qualifies as a principle of subjective wrongness relative to Q is not, of course, to say that it is an acceptable principle of this sort, or that, if acceptable, it ranks high in the hierarchy of appropriate principles of subjective wrongness relative to Q. It is only to say that principle P *should be understood and evaluated* as a candidate principle of subjective wrongness relative to Q.

What does Definition (5), together with principle Q, imply for a case in which the agent believes that act A would definitely involve killing an innocent person? In such a case, Q can serve as a principle of subjective wrongness relative to itself, prescribing the wrongness of A, given that the agent believes of A that it has a wrong-making feature stipulated by Q. Thus, Q can be a principle of subjective wrongness relative to itself when the agent has sufficiently rich non-normative beliefs to apply Q itself, whether his beliefs are correct or incorrect.[53]

[52] Note that there may be cases in which an agent has "mixed" types of beliefs. For example, the agent might believe that he has several options (e.g., A, B, and C), and might be certain that A has a wrong-making feature according to Q, but uncertain whether B or C has right-making or wrong-making features. Definition (5) needs to be revised to accommodate such cases more cleanly.

[53] Strictly speaking, it is not principle Q itself ("A is right if and only if A has F") that serves as the principle of subjective rightness, but a version of this principle stated in terms of "if" rather than "if and only if." This change is necessary to accommodate the fact that there may be more than one principle of subjective rightness. Note that Definition (5) leaves open whether the most appropriate principle of subjective rightness for an agent who has

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

Let us consider some of the implications of Definition (5). First, it implies, as I have argued it must, that one cannot simply examine the right-making features identified by a normative principle in order to ascertain that it is a principle of subjective rather than objective rightness. (Henceforward, for simplicity of exposition, I shall focus on principles of subjective *rightness*, and let the reader infer parallel statements about subjective wrongness.) One has to know whether the principle is part of a larger moral theory in which it plays the role of prescribing choices for agents who are mistaken or uncertain about which act the governing principle of objective rightness prescribes.[54]

Second, like Definition (4), it identifies the agent's beliefs as the basis (along with the governing principle of objective rightness) for the action's subjective evaluative status.[55] Definition (5) specifies that the beliefs in

sufficiently rich beliefs to apply Q itself is principle Q itself (e.g., "A is right if A has F") or a "subjectivized" version of Q that includes overt reference to the agent's beliefs (e.g., "A is right if the agent believes that A has F"). This means that Definition (5)'s clause "relative to principle Q and relative to the agent's non-normative beliefs" can be satisfied in either of two ways: the agent's beliefs can figure as part of the subjectively right-making features of the action stipulated by the principle (as is true in the subjectivized version of Q), or the agent's beliefs can figure as part of the conditions that make it appropriate to evaluate an action by a principle that specifies subjectively right-making characteristics that themselves involve no reference to the agent's beliefs. By virtue of this clause in Definition (5), every acceptable principle of subjective rightness will evaluate actions relative to the agent's beliefs.

[54] I have argued above that one cannot determine that a normative principle is a principle of subjective rightness just by ascertaining that the right-making features it identifies refer to the agent's beliefs (since some principles of objective rightness also identify right-making features that refer to the agent's beliefs). In parallel, we can now note that it is not possible to infer that a principle of subjective rightness *must* identify right-making features that refer to the agent's beliefs. If a principle of objective rightness Q can serve as a principle of subjective rightness relative to itself in a case in which the agent believes of some act that it has the right-making feature identified by Q (and this feature does not refer to the agent's beliefs), then Q, in its guise as a principle of subjective rightness, does not identify right-making features that refer to beliefs. (See the previous note.) We also know this from theorists who argue that the best principles of subjective rightness for act-utilitarianism may be the rules of common-sense morality, which have no reference to the agent's beliefs. See the discussion below under the third implication of Definition (5).

Note also that a given normative principle might have unique features that make it an appropriate principle of subjective rightness for a single principle of objective rightness. Other normative principles may be appropriate for many principles of objective rightness.

[55] Since, according to Definition (5), a principle of subjective rightness P prescribes actions relative to Q and relative to the agent's non-normative beliefs, the agent's beliefs form part of the *basis* for the subjective moral status of the agent's actions. This is true whether or not the principle of subjective rightness overtly stipulates that the agent's beliefs are part of the subjective-rightness-making features of the actions.

There is a question whether we should make subjective rightness rest on the agent's beliefs, or on all the agent's doxastic states, or on the agent's doxastic states together with relevant sub-doxastic states. We should certainly include the agent's *credences*—his degrees of belief in something. (Note that the line between "believing P" and "having credence C (very high, but less than 1.0) in P" is not a clean one, and, hence, the line between what it is best to choose in light of one's mistaken beliefs, and what it is best to do in light of one's uncertainties, may not be clean either.) We should probably include the agent's suspension of belief about some issues. But what about his unconscious or merely latent "stored" beliefs? I suspect these should not be included, since the agent may have no access to them,

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

question are the agent's *non-normative* beliefs. This enables us to deal correctly with cases in which the agent has beliefs about the objective or subjective rightness of his action, but these normative beliefs do not relate appropriately to his non-normative beliefs. For example, suppose an agent Ralph believes that his pulling the trigger has a probability of .25 of killing an innocent person—and also believes that his pulling the trigger is subjectively right. According to the principle P just described, Ralph's act is subjectively wrong relative to principle Q, even though he believes it to be subjectively right.[56] Once we note that the action an agent believes to be subjectively right may be different from the action that is actually subjectively right, the question arises which of these actions is the one most relevant for *blameworthiness.* Is Ralph blameworthy for doing what he believes to be subjectively right? In the normal case, it appears to me that agents' blameworthiness depends on what they *believe* to be subjectively right or wrong, not on what is *actually* subjectively right or wrong for them. On this view, Ralph is not to blame for pulling the trigger.[57]

and by hypothesis is not aware of them at the time of decision. Thus, the agent is not in a position to consciously guide his decision in light of these unconscious beliefs. However, further work on this issue is needed. If such unconscious stored beliefs play a causal role in agents' decision-making, it is less plausible to deny them a role in what is subjectively right for the agent. (For example, the agent may not have a conscious belief that the floor under his feet is solid, but this unconscious belief may play a causal role in his decision to step forward.)

It would be natural to think that Definition (5) should be phrased in terms of the agent's non-normative beliefs *about her action.* However, some facts that are taken by many moral codes to be relevant to an action's moral status may not be conceptualized by agents as facts about the action, so it seems best not to restrict the content of the agent's non-normative beliefs any further.

[56] What should be said about a case such as the following? Suppose the best principle of subjective rightness prescribes the act that, according to the agent's beliefs, would maximize expected value. Let us stipulate that Sue, in *Strong Medicine* (described in Section II.C–D), believes the facts described in the middle two columns of table 2, but lacks any beliefs about the facts stated in the right-most column (which describes the expected values of her options). So Sue has no belief of any action that it would maximize expected value, although the fact that giving Ron drug X would maximize expected value is entailed by her other non-normative beliefs.

I believe adherence to the Guidance Adequacy Criterion implies that we should interpret Definition (5) *not* to imply in such a case that Sue's giving Ron drug X would be subjectively right—since Sue herself does not believe of this act that it would maximize expected value. Although the contents of Sue's beliefs may entail that giving Ron drug X would maximize expected value, nonetheless she herself does not see this, since she has not derived the logical implications of her own beliefs. Perhaps in the next moment she will derive these implications. Definition (5) implies that it would *then* be subjectively right for her to give Ron drug X. The situation at the earlier time is a case in which the logical link between the contents of beliefs Sue does have and the content of the belief that would enable her to apply a given principle of subjective rightness is short and direct, so one may balk at refusing to say that giving Ron drug X would be subjectively right for Sue. However, there are other cases in which the link—although just as tight—is distant and obscure, and we are hardly surprised that the agent does not observe this link. In both cases, since we are focusing on what it is subjectively right for the agent to choose at t_i, we need to focus on what her actual beliefs at t_i would support.

[57] However, this matter is complicated. In certain pathological cases, where the agent adheres to an erroneous ethical theory, his action in accord with the absolutely subjective

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

Thus, there is a connection between subjective rightness and blameworthiness, but it is less direct than we might have supposed.

Third, Definition (5) allows room for principles of subjective rightness whose right-making features do not "match" the right-making features of the underlying principle of objective rightness. For example, the principles of subjective rightness may evaluate actions in terms of their probabilistic features, even though the governing principle of objective rightness evaluates actions in terms of their nonprobabilistic features. On some views, the lack of match could be even more extreme: for example, some act-utilitarians hold that a principle such as "It is wrong to kill an innocent person" is an appropriate principle of subjective wrongness relative to act-utilitarianism, since people are more likely to have beliefs, and indeed true beliefs, about whether their proposed action would involve killing an innocent person than they are to have beliefs about whether their action would maximize utility.[58] It is sometimes held that what makes a principle of subjective rightness appropriate to an underlying principle of objective rightness is the actual pattern of actions that agents would (or would likely) perform if they tried to follow the principle of subjective rightness.[59] This view about what justifies principles of subjective rightness implies that, so long as agents are sometimes mistaken about what objective right-making features actions have, there will be nonmatching objective and subjective right-making features.

Fourth, Definition (5) includes a clause specifying that the action is prescribed as relative to the agent's non-normative beliefs, *including her beliefs about which acts are possible for her.* This feature allows for cases in which the agent is physically unable to perform some act, but because she is unaware of this, her non-normative beliefs entail that this act would be best among all her alternatives. Thus, an agent Rachel might believe it

right-making characteristics may be blameless, even though he himself views his action as wrong and blameworthy. See Jonathan Bennett, "The Conscience of Huckleberry Finn," *Philosophy* 49, no. 188 (April 1974): 123–34. Moreover, since an agent can be criticized for performing an action that he believes to be subjectively right, but performs for the "wrong reason" (e.g., not because it is subjectively right but because it will harm his enemy), the tie cannot be as close as the text suggests. Note also, as Preston Greene points out, that the luminosity-of-beliefs problem also crops up in connection with such a definition of blameworthiness.

[58] See, for example, John Stuart Mill, *Utilitarianism,* chapter II; Sidgwick, *The Methods of Ethics,* chapters III, IV, and V; Smart, "An Outline of a System of Utilitarian Ethics," section 7; R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford: Clarendon Press, 1981), esp. section I.3 ("The Archangel and the Prole"); Shaw, *Contemporary Ethics,* 145–50; and perhaps Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," in Peter Railton, ed., *Facts, Values, and Norms* (Cambridge: Cambridge University Press, 2003): 165–68. For relevant contemporary discussion in psychology, see Gerd Gigerenzer, Peter M. Todd, and the ABC Research Group, *Simple Heuristics That Make Us Smart* (New York: Oxford University Press, 1999).

[59] This is a common (but not the only) account of what makes a principle of subjective rightness appropriate to an underlying principle of objective rightness.

would be best for her to turn the car ignition key, only to discover after she tries that she has suffered a stroke and cannot move her arm. In terms of her beliefs at the time of choice, turning the ignition key would be prescribed, and we want to recognize this fact, and give her credit (if there is any blame in question) for making the best choice, even though it turns out that this act was not possible for her.[60] This feature also allows for cases in which the agent does not believe of some action that it is physically possible for her, although in fact it is possible. Such an action might be objectively right (or wrong) according to Q, but it will have no subjective moral status.

Fifth, Definition (5) does not provide a *substantive account* of the content of principles of subjective rightness. It does not tell us, for example, that an agent who is uncertain which action would maximize utility would be subjectively right to choose the act that he believes would maximize the expectation of utility. But it is not the job of a *definition* of subjective rightness to provide such a substantive account (despite the fact that some of the definitions we examined earlier attempt to do this). The job of the definition is to provide an understanding of the concept of subjective moral status. Once we have that understanding, including a grip on the six criteria advanced in Section II.A of this essay for evaluating proposed definitions, we can proceed to find and evaluate substantive principles that will serve this role.

Sixth, it appears that Definition (5) satisfies our six criteria, or comes as close as possible. Because it bases subjective rightness on the agent's beliefs, principles that accord with it can recommend actions that strike us as reasonable or wise for the agent to choose, given his (possibly faulty) grasp of the situation. Thus, it satisfies the Normative Adequacy Criterion. Because the definition permits multiple principles of subjective rightness to augment any governing principle of objective rightness, each principle of objective rightness can be supplemented with a broad array of principles of subjective rightness designed to assess the status of every action assigned objective moral status—and, indeed, because principles satisfying Definition (5) assess the status of actions that are not possible for the agent to perform, it may ascribe subjective status to actions that cannot have any objective status. Thus, Definition (5) satisfies the Domain Adequacy Criterion.[61] However, the fact that an agent may be mistaken

[60] This will be relevant to discussions of free will and moral responsibility when the agent could do no other than what she does, as in "Frankfurt-style" cases, originally described by Harry Frankfurt in "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66, no. 23 (December 4, 1969): 829–33.

See Graham, "'Ought' Does *Not* Imply 'Can'," 4, for discussion of the fact that an act may be subjectively right even though the agent cannot perform it (although Graham dismisses the need for a concept of subjective rightness).

[61] Possibly there will be agents whose belief sets, or mental capacities, are so impoverished that *no* principle of subjective rightness can assess which action would be best for them. This, however, is a not a problem reflecting any inadequacy in Definition (5).

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

or uncertain about what her (relevant) beliefs are raises the question whether there will be cases in which an agent (such as Allison in *Learning Disability I* or *II*) cannot derive any guidance from an appropriate principle of subjective rightness. If such cases exist, then we would have to conclude that Definition (5) fails to fully satisfy the Guidance Adequacy Criterion. Since we cannot answer this question until we have seen how the notion of "the subjectively right act" should be defined, I will place this question temporarily on hold. The Relation to Blameworthiness Criterion seems to be satisfied by Definition (5), since it is reasonable to say that (in most cases) agents ought to guide their decisions by reference to what they believe to be objectively right, or in the case of cognitive impediments, by reference to what they believe to be subjectively right as characterized by Definition (5). An agent who decides to do what he believes to be either objectively or subjectively right is normally not blameworthy for his choice. Furthermore, given its generous breadth, Definition (5) appears to be compatible with the full range of plausible theories of objective moral status, and thus appears to satisfy the Normative Compatibility Criterion. Finally, Definition (5) provides some illumination about why subjectively right acts are reasonable or wise for agents to perform. Given their mistakes or uncertainty about the facts directly relevant to their principle of objective rightness, their need to make a decision, the importance of their being able to exercise moral autonomy through their decisions, and the dependence of blameworthiness on an agent's psychological states, these agents' best recourse is to guide their actions by principles that recommend actions in light of the agents' actual beliefs. The hierarchy of principles of subjective rightness provides normatively appropriate guidance. Thus, Definition (5) appears to satisfy the Explanatory Adequacy Criterion. Unlike the previous contenders we have surveyed, Definition (5) appears to be a successful characterization of what makes a normative principle a principle of subjective rightness.[62] However, the extent to which it satisfies the Guidance Adequacy Criterion remains to be determined.

[62] Definition (5), like some of the others we have reviewed, opens the question whether "subjective rightness" should be restricted, as most discussions have restricted it, to the moral status of an action relative to the *agent's* beliefs at the time of choice. Advisors and onlookers may also have beliefs in virtue of which they appraise the agent's action (or prospective action). The agent himself may have different beliefs at different times (both before and after the action) relative to which the action can be appraised. The agent may gradually gain more information in the run-up to the action, in virtue of which its "subjective" status changes; and he may gain more information after having acted, in virtue of which the action's "subjective" status may change and he may regret having chosen it. Given the importance of these additional assessments, it would be both possible and perhaps useful to broaden the definition of "subjective rightness" so that it is relative to any given set of beliefs-at-a-time. However, for purposes of this essay I will leave subjective status as defined in terms of the agent's beliefs (implicitly) at the time of choice.

V. The Subjectively Right Act

Given Definition (5), which defines when a normative *principle* is a principle of subjective rightness/wrongness, we can now provide a definition of an *act's* being subjectively right/wrong:

> *Definition (6):*
> (1) Act A (which would be performed at time t_j) is *subjectively right* at t_i relative to principle of objective status Q just in case A is an act prescribed by the highest principle of subjective rightness relative to Q that the agent is able to use as an internal guide at t_i; and
> (2) Act A (which would be performed at time t_j) is *subjectively wrong* at t_i relative to principle of objective status Q just in case A is an act proscribed by the highest principle of subjective wrongness relative to Q that the agent is able to use as an internal guide at t_i.[63]

Definition (6) only states what makes an act count as subjectively right or wrong relative to some principle of objective status or other. However, we may want to know whether the act is subjectively right or wrong relative to the *correct* principle of objective status. To capture this idea, we can define the concept of *absolutely* subjectively right/wrong actions:

> *Definition (7):*
> (1) Act A (which would be performed at time t_j) is *absolutely subjectively right* at t_i just in case A is an act prescribed by the highest principle of subjective rightness (relative to the correct principle of objective rightness) that the agent is able to use as an internal guide at t_i; and
> (2) Act A (which would be performed at time t_j) is *absolutely subjectively wrong* at t_i just in case A is an act proscribed by the highest principle of subjective wrongness (relative to the correct principle of objective wrongness) that the agent is able to use as an internal guide at t_i.

Both Definitions (6) and (7) utilize the concept of an agent's being "able to use" a given principle of subjective status as an internal guide. The basic idea, articulated in Section II.A of this essay, is that the agent can derive a prescription or proscription for an action from the principle. But in an obvious sense an agent often "can derive" a prescription from a

[63] Note that an act may be subjectively right at t_i (because it is prescribed by the highest principle of subjective rightness the agent can use at t_i) even though the agent does not ask himself at t_i the question of whether to perform the action, or whether it would be subjectively right to perform the action.

Definition (6) would have to be further developed to handle cases (such as the Regan-type case, described in note 26) in which the agent has *mixed* information about his various possible options—for example, having beliefs about what the expected value of some acts would be, but not having any beliefs about the expected value of other acts.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

principle, even if the agent cannot derive the prescription for any action under what Eugene Bales calls "an immediately helpful description."[64] Some descriptions may accurately pick out an action, but not in a manner that enables the agent to identify it in such a way as to perform it if he wants to. Such descriptions are "unhelpful." Thus, someone who has no idea what the consequences would be of his various alternatives can still derive a "prescription" from act-utilitarianism—he can derive the prescription "Perform the act that would maximize utility." The description "act that would maximize utility" picks out a unique act, but this is no help if he cannot identify which act this is in terms that would enable him to perform it in the way that describing the act as "Tell the employees to use the stairwell" enables one of the security guards to perform this act. To get around this problem, we need a somewhat complicated definition, as follows (here, again, I shall focus just on prescriptive principles):[65]

> *Definition (8):*
> An agent S is able at t_i to use normative principle X as an internal guide to decide at t_i what to do at t_j just in case (1) there is some (perhaps complex) feature F such that X prescribes actions that have feature F, in virtue of their having F; (2) S believes at t_i of some act-type A that S could perform A (in the epistemic sense) at t_j;[66] (3) S believes at t_i that if she performed A at t_j, her action would have feature F; and (4) if S wanted at t_i to derive a prescription from

[64] Bales, "Act-Utilitarianism," 261.

[65] There is a highly developed literature on rule-following that focuses on questions somewhat distinct from those at issue in this essay. See, for example, Peter Railton, "Normative Guidance," in Russ Shafer-Landau, ed., *Oxford Studies in Metaethics*, vol. 1 (Oxford: Oxford University Press, 2006), 3–34.

[66] That is, slightly revising Alvin Goldman's definition of "ability to perform an act" in the epistemic sense, S believes (doubtless expressed in her own concepts) that

(1) There is an act-type A* which S truly believes at t_i to be a basic act-type for her at t_j;
(2) S truly believes that she is (or will be) in standard conditions with respect to A* at t_j; and
(3) either
  (a) S truly believes that A* = A, or
  (b) S truly believes that there is a set of conditions C* obtaining at t_j such that her doing A* would generate her doing A at t_j.

See Alvin I. Goldman, *A Theory of Human Action* (Englewood Cliffs, NJ: Prentice-Hall, 1970), 203. Roughly speaking, a person is in standard conditions with respect to an act property just in case (a) there are no external physical constraints making it physically impossible for the person to exemplify the property, and (b) if the property involves a change into some state Z, then the person is not already in Z. See ibid., 64–65. Note that on Definition (8) the agent believes that she truly believes there is a basic act-type for her, etc., but she may be wrong about what she believes and whether her belief is true.

Further complications would have to be introduced to deal with cases in which the agent is uncertain whether some act is one she can actually perform, and to deal with deviant causal chain cases.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

principle X at t_i for an act performable at t_j, S would derive a prescription for A in virtue of her belief that it has feature F.[67]

For example, Rachel (the unwitting stroke victim) is able to use the normative principle "Maximize utility" to make a choice, since (i) this principle prescribes actions having the feature that they will maximize utility; (ii) Rachel believes that she can turn the ignition key; (iii) she believes that turning the ignition key would maximize utility; and (iv) if she wanted to derive a prescription from the principle "Maximize utility," she would derive a prescription to turn the ignition key in virtue of its being the act that would maximize utility.

Note several implications of Definition (8). First, while S believes she can perform A, it may not be true that she can. Second, while S believes that if she performed A, her action would have feature F, in reality her performing A may not have F. Third, the time at which the choice would take place is not necessarily identical with the time at which the act would take place; one can choose now to perform an act later on (although typically one has to reaffirm this choice when the time for action comes). Fourth, S may believe that there are several acts performable at t_j that have feature F; for S to be able to use X to make a choice, all that is necessary is that S would derive a prescription for *one* of these acts.

Definitions (6) and (7) also utilize the notion of a principle of subjective rightness being "the highest" principle of subjective rightness relative to some principle of objective rightness. As I have noted above, to accommodate the great variation in the kinds of beliefs agents have when they must make moral decisions, we need a rough hierarchy of principles of subjective rightness that are appropriate for a given principle of objective rightness. Thus, agent S's beliefs might make it possible for her either to use principle P_1 (advising her to maximize expected utility) or to use principle P_2 (advising her to minimize the maximum loss of utility). Both of these principles may have a place in the hierarchy of principles of subjective rightness appropriate for the act-utilitarian principle of objective rightness. But if S is able to use either one, P_1 is arguably higher in the hierarchy than P_2, and the act prescribed by P_1 is subjectively right relative to act-utilitarianism.

Using these new tools, let us now ask whether or not Definition (5) licenses principles of subjective rightness that satisfy the Guidance Adequacy Criterion, which requires that a definition of subjective rightness should endorse principles of subjective rightness that agents are able to use as an internal guide for decision in every situation in which they find themselves, even though an agent may be mistaken or uncertain about which actions have the features that would make them objectively right in that situation.

[67] One would want variants on this for actions that are forbidden, but since our main focus is on an agent's deciding what to do (not just what not to do), in the interests of shorter exposition I will omit these variants.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

Since Definition (5) identifies a principle as a principle of subjective rightness, relative to some governing principle of objective rightness, based on the beliefs of the agent, and since agents are typically better informed about their beliefs than they are about the circumstances and consequences of their actions, it appears that Definition (5) should have little problem meeting the Guidance Adequacy Criterion. But we have seen that agents are sometimes mistaken or uncertain about their beliefs. How does Definition (5) deal with these situations?

To see this, let us consider the following moral theory (MT-1) and its implications in a case in which the agent is mistaken about her own non-normative beliefs. MT-1 is comprehensive, in the sense that it includes not only a principle of objective rightness, but also principles of subjective rightness, a rank-ordering of these principles, and a statement of when it deems an action to be subjectively right.

MT-1

Principle of objective rightness:
Q: An act X is objectively obligatory if and only if X maximizes value.

Principles of subjective rightness:
P: An act Y is a candidate for being subjectively obligatory if Y would maximize value.[68]
R: An act Z is a candidate for being subjectively obligatory if Z would maximize the minimum value.

The subjectively right act:
(a) Principle P is higher than principle R; and
(b) An act W is subjectively obligatory if and only if W is prescribed by the highest principle of subjective rightness listed above that the agent is able to use as an internal guide.

Consider how MT-1 applies in the following (abstract) case in which the agent (S) is mistaken about her own beliefs:

CASE 1

(a) Agent S believes MT-1 is the correct moral theory.
(b) S believes of act A that it would maximize value.

[68] Note that the principles of subjective rightness are phrased as sufficient conditions ("... if ...") rather than as necessary and sufficient conditions ("... if and only if ..."). This phrasing is needed to accommodate the fact that there may be many principles of subjective rightness, so each can only offer a sufficient (but not necessary) condition for an act's being a candidate for being subjectively right.

(c) S does not believe that she believes of any act that it would maximize value (this is S's mistake about her beliefs).
(d) S believes of act B that it would maximize minimum value.
(e) S believes that she believes of act B that it would maximize minimum value.
(f) Act A would maximize value.
(g) Act B would maximize minimum value.

Taking MT-1, the facts in Case 1, and our definitions of what it is for a principle to be internally usable (Definition [8]) and of what it is for an act to be subjectively right relative to a principle of objective rightness (Definition [6]), we can infer the following:

(1) According to Definition (8), on the straightforward version of the psychology in this case, principle P is *not* usable by S, because it is false that if S wanted to derive a prescription from P, she would do so. (She would fail to derive a prescription from P because she does not believe that she believes of any act that it would maximize value.)
(2) S would believe that P is not usable by her.
(3) According to Definition (8), principle R *is* usable by S, since she believes of act B that it would maximize the minimum value, and it is true that if she wanted to derive a prescription from R, she would do so.
(4) Thus, R is the highest usable principle of subjective rightness for S.
(5) In light of her information, S is in a position to conclude that R is the highest usable principle of subjective rightness for her.
(6) Hence, S is in a position to conclude that act B is subjectively right, since she is in a position to conclude that B is prescribed by the highest usable principle of subjective rightness relative to principle Q.
(7) Act A is not subjectively right, because even though it is prescribed by principle P (the highest principle of subjective rightness relative to Q), principle P is not usable by S.
(8) Act B is in fact the subjectively right act for S relative to principle Q, since it is prescribed by the highest usable principle of subjective rightness relative to Q.[69]

[69] In point (1) of this list of eight points, we construed the case as one in which principle P is not usable by S, since she does not believe that she believes of any act that it would maximize value. But, alternatively, the psychology of the case could be such that P *is* usable by S, since, given that S actually does believe of act A that it would maximize value, she might (to her surprise) derive a prescription for A from P. On this construal, the case would turn out as follows:

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

Thus, MT-1, founded on Definition (5), provides an internally usable decision guide for this agent, even though she is mistaken about her relevant beliefs.[70] Case 1 serves to reassure us that an agent's mistakes *about her beliefs* will not prevent her from using principles of subjective rightness as internal decision guides.

Suppose that in Case 1, S is mistaken in believing that act B would maximize minimum value; in fact, some third act C has this characteristic. Then act B would not actually be subjectively right for S; instead, act C would. This case shows us that principles of subjective rightness, even when they are usable as *internal* decision guides, are not necessarily usable as *external* decision guides. We suspected from the beginning that even principles of subjective rightness would not succeed as external decision guides, and this case confirms this suspicion. Just as there can be a gap between what is objectively right for an agent and what she believes is objectively right, there can be a gap between what is subjectively right and what she believes is subjectively right. Invoking the concept of subjective rightness does not ensure that agents are infallible when they seek to perform the wisest action. In our case, this feature does not depend on the agent's mistakes about her own beliefs. It could crop up in any case in which the agent is mistaken in believing that an action has a subjective-right-making feature that it lacks.

The analysis of MT-1's usability can be duplicated for cases in which an agent is *uncertain* about her own beliefs—for example, the agent actually believes that act A would maximize value, but is uncertain whether she believes that A would maximize value. In these cases, too, the agent is able to derive internal guidance for what to do.[71] Thus, it appears that mistakes

---

(1′) Principle P is usable by S, since she believes of act A that it would maximize value, and if she wanted to derive a prescription from P she would do so, in virtue of this belief.
(2′) Thus, Principle P is the highest usable principle of subjective rightness for S.
(3′) In light of her information, S is in a position to conclude that P is the highest principle of subjective rightness usable by her.
(4′) Hence, S is in a position to conclude that act A is subjectively right, since she is in a position to conclude that A is prescribed by the highest usable principle of subjective rightness relative to Q.
(5′) Act A is prescribed by the highest usable principle of subjective rightness, and so is subjectively right.

On this alternative construal of this case, S is also able to use one of the principles of subjective rightness for Q as an internal decision guide.

Note that if 3′ ("S is in a position to conclude that P is the highest principle of subjective rightness usable by her") is false, then S would mistakenly conclude that A is not subjectively right.

[70] Note that S could have mistaken normative beliefs (she might not believe MT-1 contains the correct principle of objective rightness, or she might mistakenly believe that principle of subjective rightness R is higher than principle P, or she might not be able to grasp any or some of these principles). These cognitive errors, too, may lead her astray in various ways. These are complications I explore in *Making Morality Work*.

[71] Similarly, the analysis can be duplicated for moral theories that are subjectivized, i.e., ones in which principles such as P explicitly refer to the agent's beliefs as grounds for the

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

or uncertainty about her non-normative beliefs do not stand in the way of an agent's finding an internally usable guide for her decision-making, appropriate to a principle of objective rightness Q, so long as Q is supplemented by a rich enough set of principles of subjective rightness (and the agent is familiar with these). Although normative ignorance or mistake may stand in her way, non-normative ignorance or mistake, even about her own beliefs, will not. Definition (5), when combined with Definitions (6) and (8), appears to license moral theories whose subordinate principles of subjective rightness will jointly meet the Guidance Adequacy Criterion.

VI. Reasonable Beliefs as the Ground for Subjective Rightness

As I have noted, many theorists hold that "subjective rightness" should be defined in terms of what it would be reasonable for the agent to believe (or what an agent would be justified in believing, etc.), rather than in terms of what the agent actually believes. Given the tools we have developed, let us consider this suggestion more fully. In light of our earlier rejection of this approach based on Definitions (1) through (4), the natural strategy for a proponent of this approach would be to offer a revised version of Definition (5) that incorporates reference to reasonable beliefs in place of (5)'s reference to actual beliefs. A version of such a definition (here stated for rightness only, in the interests of brevity) could be stated as follows:

Definition (5):*
If Q is a principle of objective moral status, and Q stipulates that F is a right-making feature of actions and that G is a wrong-making feature of actions, *then*

(1) A normative principle P is a principle of *subjective rightness* relative to principle Q just in case, for any agent S, either of the following is true:

(A) *if* it would be reasonable for agent S to believe of some act A that A is possible for him to perform and that A has feature F, *then* principle P prescribes A, relative to principle Q and relative to the non-normative beliefs that it would be reasonable for S to have about A; or
(B) *if* (i) it would be reasonable for agent S to believe of some act A that A may be possible for him to perform, *and if* (ii) it would be reasonable for S to be uncertain whether any act

subjective status of the action, as in "An act Y is a candidate for being subjectively obligatory if the agent *believes that* Y would maximize value." See note 53 for discussion of "subjectivizing" a moral principle.

available to him has feature F, and if so, which act does have F, *then* principle P prescribes A relative to principle Q and relative to the non-normative beliefs it would be reasonable for S to have about A.[72]

The standard principles of subjective rightness licensed by Definition (5)* would refer to the beliefs it would be reasonable for an agent to have. For example, such a principle might state that an act would be subjectively right just in case it would be reasonable for the agent to believe that it would maximize expected value. Unfortunately for this approach, it is clear that agents frequently have no beliefs about what it would be reasonable for them to believe, or are uncertain or mistaken about what it would be reasonable for them to believe. Hard-pressed decision-making agents typically do not ask themselves what it would be reasonable for them to believe. And even an agent who does ask herself this, and realizes that she should have investigated further (or deliberated more carefully or longer) before making a decision, and who therefore believes that her present beliefs may not be reasonable, may have little idea what beliefs about the facts she *would* have had if she had investigated or deliberated further.

What does this imply about the acceptability of Definition (5)*? To see this, consider the following moral theory (MT-1*) and its implications. MT-1* is modeled on MT-1, but substitutes "it would be reasonable for the agent to believe" for "the agent believes."

*MT-1**

Principle of objective rightness:
Q: An act X is objectively obligatory if and only if X maximizes value.

Principles of subjective rightness:
*P**: An act Y is a candidate for being subjectively obligatory if it would be reasonable for the agent to believe that Y would maximize value.
*R**: An act Z is a candidate for being subjectively obligatory if it would be reasonable for the agent to believe that Z would maximize the minimum value.

The subjectively right act:
(a) Principle P* is higher than principle R*; and
(b) An act W is subjectively obligatory if and only if W is prescribed by the highest principle of subjective rightness listed above that the agent is able to use as an internal guide.

[72] Note that a version of Definition (5) phrased in terms of the beliefs S actually has that *are reasonable* would not be tenable, since many agents would have no reasonable beliefs relevant to the choice they must make, and yet still need guidance in making that choice.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

Now consider a case in which the agent does not have the relevant beliefs about what it would be reasonable for her to believe.

CASE 2

(a) Agent S believes that MT-1* is the correct moral theory.
(b) It would be reasonable for S to believe of act A that it would maximize value, and reasonable to believe of act B that it would maximize minimum value.
(c) S does not believe that it would be reasonable for her to believe of any act that it would maximize value, or would maximize minimum value.
(d) Neither principle P* nor principle R* is usable by S, because it is false that if S wanted to derive a prescription from P* or from R*, she would do so. (She would fail to derive a prescription from either of these principles because she does not believe of any act that it would be reasonable for her to believe of that act that it would maximize value, or maximize minimum value.)
(e) Since neither P* nor R* is usable by S, there is no act which is subjectively right for S to perform.
(f) S would not conclude about any act that it is subjectively right for her to perform it.

Thus, application of MT-1* to Case 2, in which S does not believe that it would be reasonable for her to believe of any act that it either maximizes value or maximizes minimum value, indicates that there is no act that is subjectively right for S, and MT-1* provides S with no usable internal decision guide.

Of course, MT-1* is a highly impoverished theory, and could be expanded by adding more principles of subjective rightness. This might help the usefulness of MT-1*, since the expanded version might include some lower-level principle of subjective rightness which would be usable even if higher-level principles are not. For example, if the expanded MT-1* includes principle T* ("An act W is a candidate for being subjectively right if it would be reasonable for the agent to believe that W might produce some positive value"), and S believes it would be reasonable to believe of act A that it might produce positive value, then T* would be usable by S as an internal guide for making decisions according to MT-1*.

But even though the lower-level principles of subjective rightness for the expanded MT-1* (such as principle T*) would not place heavy demands on the agent's beliefs about what it would be reasonable for her to believe, nonetheless there will be many cases in which the agent must decide here and now what to do, and in which—because she hasn't asked herself the question—she has no beliefs about what it would be reasonable for her to believe. In such cases, even an expanded MT-1* would not provide any

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

decision guide for the agent—even though the agent may well have beliefs about various features her actions have, and so would be able to use MT-1 to guide her decision.[73]

Thus, since agents often lack beliefs about what it would be reasonable for them to believe about their various options, Definition (5)* licenses moral theories that are usable in a significantly smaller range of cases than the moral theories that are licensed by Definition (5). Adoption of Definition (5)*, as opposed to Definition (5), would result in numbers of agents who lack any usable principle of subjective rightness at all. I conclude that we should reject Definition (5)* on the grounds that it cannot provide sufficiently widely usable decision guides.[74] Even an agent who has given no thought to what it would be reasonable for her to believe, or has no idea which belief would be reasonable, still has to make a decision, and her moral theory should enable her to do so.

[73] For every moral theory, there may be a "bottom-level" principle of subjective rightness—the lowest principle in the hierarchy, to be used when the agent completely lacks any relevant information about his prospective acts. It is plausible that, for MT-1* (or any moral theory), the bottom-level principle should designate as morally permissible any act the agent can perform, since, by hypothesis, the agent has no way to rule out any act as inconsistent with the values of the principle of objective rightness. Thus, the bottom-level principle of subjective rightness for MT-1* would be "An act W is a candidate for being subjectively permissible if W is an act that it would be reasonable for the agent to believe he can perform." Such a principle makes very limited cognitive demands on an agent. Nonetheless, it makes more demands than the parallel principle for MT-1 ("An act W is a candidate for being subjectively permissible if W is an act that the agent believes he can perform"), since it still requires that the agent have beliefs about what it is reasonable for him to believe—and many agents may not have such beliefs, either because they are not thinking about what it is reasonable for them to believe, or because they are uncertain what it is reasonable for them to believe. Thus, even when it is augmented by such bottom-level principles, MT-1* is less widely usable than MT-1.

[74] One of the major arguments in favor of defining subjective rightness in terms of beliefs that it would be reasonable to have, rather than in terms of actual beliefs, is that "reasonable beliefs" rather than "actual beliefs" are arguably the beliefs most relevant to the agent's blameworthiness. This position on blameworthiness is itself controversial. I would argue that it is incorrect: while it is true that an agent may be blameworthy for not making the inquiries she could and should have made (or for not drawing the correct conclusions from her evidence), it does not follow from this that she is blameworthy for making the choice that appears best in light of the directly relevant beliefs she actually has at the time of decision. The role of principles of subjective rightness is to provide her with the guidance she needs and can use at the time she must make her decision, not the guidance that a better agent could use. For further discussion, see my "Culpable Ignorance," *The Philosophical Review* 92, no. 4 (October 1983): 543–71. But even a theorist who holds that the blameworthiness of an agent depends on the beliefs it would be reasonable for her to have (as opposed to those she actually has) should still accept the original Definition (5) of subjective rightness, since it—but not Definition (5)*—provides autonomy to agents seeking to guide their decisions by reference to their potential acts' moral value. This theorist can then define "blameworthiness" in terms, not directly of the agent's performing what she believes to be the objectively or subjectively right act, but rather in terms of the agent's performing what a reasonable agent would have believed to be the objectively or subjectively right act. This conception needs further refinement, however, since surely an agent may blamelessly choose an act while mistakenly (but perhaps reasonably) believing it to be what a reasonable person would have believed to be subjectively wrong.

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press

VII. Conclusion

The concept of subjective rightness was originally introduced to enable us to deal with two issues: (1) the paradoxical tension between (a) what is best for an agent to do in light of the actual circumstances in which she acts and (b) what is wisest for her to do in light of her mistaken or uncertain beliefs about her circumstances; and (2) the need to provide moral guidance to an agent who may be uncertain about the circumstances in which she acts, and hence is unable to use her principle of objective rightness directly in deciding what to do. Surprisingly, there have been relatively few attempts to provide a clear and detailed analysis of the concept of subjective rightness. In this essay, I have described criteria of adequacy for any successful definition of subjective rightness, canvassed the major existing strategies for defining this notion, and rejected each of them as inadequate. I then argued we must take a different approach to the problem, focusing on defining *principles* of subjective rightness rather than subjectively right *acts.* I proposed Definition (5), which captures the crucial insight that a normative principle can be characterized as a principle of subjective rightness only relative to a governing principle of objective rightness. Along the route, I have argued that the concept of subjective rightness should be defined by reference to the agent's actual beliefs, rather than by reference to the beliefs it would be reasonable for an agent in her position to have. Definition (5) provides a solid framework for addressing our two issues: it enables us to dissolve the tension of issue (1) by distinguishing what an agent ought objectively to do from what she ought subjectively to do, and it enables us to address issue (2) by using principles of subjective rightness to provide moral guidance to agents who are uncertain about the circumstances or consequences of their actions. Armed with Definition (5), we can recognize that each moral theory must include a multiplicity of principles of subjective rightness to address the epistemic situations of the full range of moral decision-makers. Definition (5) places us in a position to evaluate and rank-order substantive principles of subjective rightness, to explore more adequately the links between subjective rightness and blameworthiness, and to assess the Remodeling proposal that principles of subjective rightness be elevated to the status of principles of objective rightness. There is much work to be done, but the groundwork has been laid.

Philosophy, Rutgers University

https://doi.org/10.1017/S0265052509990161 Published online by Cambridge University Press