respond unreflectively to a specific situation. Their responses draw on a conventional moral code and strong feelings that have grown up around the code. They say things like "Don't break promises to your friends" and "Tell your parents the truth." But when they are called on to defend their judgments, they appeal to general principles—in particular, they try to show that the conventional moral code is the best way to promote overall happiness. So there is variation in morality across groups and societies. But the variation is explained, as the justifications suggest, by beliefs about how to apply the principle of utility in different situations.

What response is available to Mackie? For an argument along these lines, aiming to show that commonsense morality is implicitly utilitarian, see Henry Sidgwick, *Methods of Ethics* (Hackett, 1980), Book 4.

3. Mackie thinks that we erroneously believe that values are objectively prescriptive. He endorses an **error theory** about values, according to which people quite generally and persistently hold a mistaken belief about values. You find error theories in other areas of philosophy as well. Error theorists about causation think that people generally believe that causality is a real connection between events (one event *makes* the other happen), though there is no such real connection; error theorists about color think that people generally assume colors to be objectively present in objects, though colors are only a matter of how we respond to objects; error theorists about scientific unobservables (say, quarks) think that people generally believe that there are unobservable objects even though talk about unobservables is simply a way of predicting observations; error theorists about aesthetics think that beauty is only in the beholder's eyes, though we believe (mistakenly) that beauty is present in the world.

   Because error theorists claim to have detected a pervasive and persistent mistake, they typically offer explanations of the roots of the error: a story about our *objectification errors* that explains how we have ended up projecting some feature of our thinking, feeling, talking, or interacting onto the world. In section 10 of "The Subjectivity of Values" (omitted from this reading), Mackie suggests some explanations of the objectification error he detects in moral philosophy and ordinary moral thought. One explanation begins with the fact that we sometimes want people to do something—say, keep a promise. Instead of saying "You should keep your promise because I want you to" or "You should keep your promise because we want you to," we just say: "You ought to keep your promise," or, more simply, "You promised." Morality thus involves "suppressing any explicit reference to demands."

   (i) Why would we want to suppress any explicit reference to demands?

   (ii) How could Mackie's explanation work if people are aware that moral claims are simply a shorthand for expressions of demands?

4. According to the (unhappily named) argument from queerness, objective values, if there were such, would need to be objects, or relations, or qualities unlike anything we are familiar with. What precisely makes objective values so unusual? Consider three answers:

   (i) Values are not in space-time.

   (ii) Values lead people who fully understand them to act.

   (iii) Values instruct us about what to do.

   Which answer does Mackie offer? Once you have settled on an interpretation of Mackie, consider whether he is right that commonsense morality assumes such odd things to exist.

R. Jay Wallace (born 1957)

Wallace is a professor of philosophy at the University of California, Berkeley, where he holds the Judy Chandler Webb Distinguished Chair for Innovative Teaching and Research. He writes on moral philosophy, with a particular focus on moral psychology and practical reasoning.

MORAL SUBJECTIVISM

Moral thought is commonly supposed to be a matter of subjective attitude, in a way that contrasts with thought about (say) mathematics or the natural world. If you judge that $12 \times 3 = 36$, or that the cat is sleeping on the bed, nobody is likely to conclude that that is just your opinion. Your thoughts seem to be about a subject matter that is prior to and independent of them, and in that respect objective. We might say that judgments of these kinds are answerable to independent facts of the matter, insofar as their correctness or incorrectness depends on how things are independently of the judgments being made.

With moral thought, by contrast, things are often taken to be otherwise. Consider the judgments that it is wrong to make insincere promises or to exploit the weak and vulnerable. It is widely believed that judgments of this kind are not answerable to a prior and independent subject matter, but are merely a matter of subjective opinion. This thought is the animating idea behind moral subjectivism. To a first approximation, subjectivism is the position that moral judgments—such as the judgment that lying promises are wrong—are not about a set of facts that are prior to and independent of them. Instead, the subjectivist maintains, they reflect the attitudes of the person who makes those judgments.[1]

1. The term "subjectivism" is sometimes used more narrowly in philosophical discussion, to refer to the view that moral judgments are about an agent's subjective states; the dispositionalist position discussed below is a subjectivist view in this more narrow sense. [Wallace's note.]

Two aspects of moral thought particularly encourage this subjectivist interpretation of it.[2] One concerns its subject matter. Moral judgments typically involve evaluative or normative concepts, as applied to persons and their actions. In moral thought we conclude that doing X would be permissible or required, right or wrong, and we judge that people are admirable or blameworthy in virtue of their character traits and the things they have done.[3] Concepts such as permissibility, rightness, wrongness, or blameworthiness, however, do not seem to correspond to any objects or properties in the natural world. Actions are not wrong, for instance, in the way the leaf of a tree might be green or oblong or bitter to the taste. We see that the eucalyptus leaf is green and oblong when we look at it, and we taste its bitterness when we put it in our mouth; these properties can affect our sense organs, in ways that make them potential objects of empirical investigation and scientific study. But the properties involved in moral thought do not in the same way seem to make a causal difference to our experiences. We don't, after all, have any special organs of perception or sensation that enable us to detect the wrongness of acts of lying.[4]

Considerations of this kind make it natural to suppose that the world is devoid of the evaluative and normative properties that moral thought apparently trades in. But if there are no evaluative and normative properties in the world, it seems to follow that moral thought cannot be understood in objective terms. It does not answer to a set of independent facts about the way things are in the world, since the world as we find it has no place for evaluative and normative objects and properties.

A second consideration that encourages the subjectivist interpretation concerns the effects of moral thought on action. One of the important ways we use moral concepts is in deliberation — the kind of systematic reflection we engage in when we attempt to get clear about what we ought to do.[5] In deliberation, we take it for granted that we could choose to act in a number of different ways (keeping a promise or breaking it, say), and we reflect on those alternatives, asking, among other things, whether they are morally permissible or required. The thoughts that figure in deliberative reflection are in this way practical in their subject matter: they are about what to do. But deliberation is practical in a very different sense as well. After reaching a conclusion about what they ought to do, those who engage in deliberation often act

2. A third aspect of moral thought that is sometimes cited in this connection is the fact of disagreement about what it is right or wrong to do. But this consideration strikes me as less significant than the other two, so I shall set it aside in what follows. [Wallace's note.]

3. Normative concepts are concepts that involve the ideas of a reason or a requirement, whereas evaluative concepts involve ideas of the good. Here I gloss over large issues about the relation between reasons and values. I also simply assume, throughout my discussion, that moral thought is a species of normative thought, concerning a special class of reasons or requirements. [Wallace's note.]

4. Note that this argument apparently also applies to mathematical thought, which similarly does not seem to be about objects and properties that we interact with causally. And yet, mathematical thought seems to be a paradigm of objectivity. Does this undermine the argument from metaphysics? [Wallace's note.]

5. Another important context in which normative thought figures is that of advice, where we reflect on the options that other people face, and try to arrive at conclusions about what they ought to do. In what follows I shall focus primarily on contexts of deliberation, in which agents reflect on their own options for action. But you should consider how the subjectivist approaches I sketch might be extended to apply to contexts of advice. [Wallace's note.]

on the verdict they have arrived at, choosing the option that deliberation has identified to be for the best.

Consider the members of a campus club who, after deliberation, decide that it is wrong to maintain their secret policy of excluding people from certain ethnic groups, even if doing so would be to their advantage. (Perhaps there are wealthy benefactors who will stop supporting the group if it becomes more inclusive in its membership.) Having arrived at this moral conclusion, the club members might adjust their policies accordingly, opening the club to people from all ethnic and cultural backgrounds, because they have come to see that that is the right thing to do. Moral thought might not have as much influence on action as we would like, but it is at least capable of moving people directly to act. It is thus practical not merely in its subject matter, but also in its effects.

This practical dimension of moral thought appears difficult to make sense of if we understand such thought in objectivist terms. The judgments that seem paradigmatically about a realm of independent objects and relations do not have this kind of influence on the will. The thought that fresh beets are available for sale in the local supermarket, for instance, does not on its own seem able to move us to action one way or another. To do so, it would need to combine with some distinct attitude on our part, such as a desire to have roast beets for dinner; if you hate beets, or are simply indifferent to them, then the true belief that you can buy some at the local supermarket will have no effect on your motivations whatsoever.[6] Moral thought, by contrast, seems capable of engaging the will directly, without the addition of attitudes that are extraneous to it. The conclusion that it is wrong to discriminate against members of certain ethnic groups, as we saw above, is already apt to move us to action by itself. It is natural to hypothesize that such conclusions must essentially involve the subject's desires or emotions, mobilizing the kinds of subjective attitudes that move us to act.

We might refer to these two lines of thought as the arguments from metaphysics and from motivation, respectively. They were both taken very seriously by David Hume, who was led by them to the subjectivist conclusion that morality "is more properly felt than judg'd of."[7] Hume meant by this that moral deliberation trades in attitudes of emotion or desire of the kind that move us to action, rather than judgments about an independent set of normative and evaluative facts. This conclusion is an extremely tempting one when we think about moral thought, and it contains at least a grain of truth. In the end, however, I don't believe that we should accept the Humean position. In support of this claim, I shall begin by considering a simple version of moral subjectivism, and then explore three different ways of refining the position. A particularly important theme will be the role of *critical reflection* in normative and moral thought: its role, that is, as a method of scrutinizing and improving our own subjective reactions. I shall argue that this is something the subjectivist cannot give an adequate account of.

Perhaps the simplest way to develop the subjectivist idea is to interpret it as a form of *expressivism*. This is a view in what is often called metaethics, the study of the meaning of the language that is used to express moral claims. Expressivists hold

6. Of course, you might have promised a friend that you would pick up some beets for him in the store. But then you will be led to act by the moral thought that you will wrong your friend if you fail to do what you promised, a thought that is also extraneous to your factual belief about the availability of the beets. [Wallace's note.]

7. David Hume, *A Treatise of Human Nature* (Oxford University Press, 1978), p. 470. [Wallace's note.]

that moral and other normative judgments are not in the business of representing a set of independent facts or relations. Their function is instead to give expression to practical attitudes of approval or disapproval, such as desires or intentions.[8] Moral language, on this approach, might be compared to the verbalizing that goes on at a football game or a rock concert, which does not even attempt to make claims about the way things are in the world, but rather gives expression to the spectators' attitudes toward the events they are observing. To say that it is wrong to exploit and mock the vulnerable, for instance, is to give voice to your disapproval of acting in this way; it expresses a desire that people should not perform actions of this kind, much as the lusty booing that takes place at the football stadium expresses the audience's disapproval of the botched play that just took place on the field.

This expressivist position does a good job of accommodating the considerations marshaled in the arguments from metaphysics and motivation. According to expressivism, moral and other normative assertions don't really say anything at all about the world, so we can make sense of such discourse without postulating any funny properties or states of affairs. The expressivist account also offers a nice explanation of the practical dimension of moral thought. If moral discourse is in the business of expressing the agent's desires, then we can immediately understand how it is that moral judgments can directly engage the will. The practical attitudes that moral discourse expresses guarantee that such motivations will be present whenever a moral judgment is endorsed.

The problem, however, is that the simple expressivist view seems to go too far in the direction of assimilating moral thought to the formation of such practical attitudes. If people can be motivated to act directly by their normative judgments, this connection can also break down. You might for instance think that it would be wrong to keep a wallet that you have found in the university library (rather than turning it in at the lost-and-found office), but give in to the temptation to keep the wallet when you realize how much money it contains. In cases of this kind, people act against their own moral judgments,[9] and the possibility of doing this suggests that moral judgments don't simply involve the expression of effective motivating attitudes.

Normative thought has an important critical dimension. It can be brought to bear on our own emotions and desires, including the motivations that lead us to act when we go astray by our own lights. This dimension of moral thought needs to be accounted for in an adequate development of the subjectivist position. The challenge is to explain how we can achieve critical distance from our motivating attitudes, within a framework that understands moral thought essentially in terms of such attitudes.

One way of responding to this challenge would be to modify the simple version of expressivism by restricting the class of subjective attitudes that moral and normative language is taken to express. In this spirit, normative discourse might

8. See, for example, Allan Gibbard, *Thinking How to Live* (Harvard University Press, 2003). [Wallace's note.]

9. The phenomenon of action against one's better judgment is often referred to as "akrasia" (from the ancient Greek) or "weakness of will." [Wallace's note.]

be supposed to give voice to our second-order attitudes, including above all the preferences we form about our first-order desires.[1] When you act against your better judgment concerning the permissibility of keeping the wallet, for example, you have a first-order desire to hang on to the money that the wallet contains. But you also form a distinct attitude about that desire, preferring that it should not prove effective in determining what you do. The subjectivist might say that it is second-order desires of this kind that it is the distinctive function of normative language to express. On the resulting picture, practical attitudes are subject to criticism by something outside themselves; but the standards for such critical assessment are fixed by further practical attitudes of the agent.[2]

A natural question to ask about this more sophisticated expressivism, however, concerns the standing of higher-order attitudes to constitute a basis for critical assessment. Suppose you form a second-order desire that your desire for money should not prevail in determining whether you keep the wallet you have found. This higher-order desire is an attitude of the same basic type as the first-order attitude that is its object; it is just another desire or preference that you are subject to. If there is a real issue about the credentials of the first-order attitude, it is hard to see how it can be resolved simply through the formation of further attitudes of the same basic kind. Won't those attitudes be prone to further iterations of skeptical undermining? You could, after all, step back from your second-order desire regarding the original temptation to keep the wallet, and call that desire into question in turn, forming a third-order desire to ignore the scruples of conscience. Nothing in the nature of your second-order preferences seems to block such critical questions from being raised about them.

The sophisticated expressivist might respond by noting that we generally don't extend the process of reflection to such extremes. We step back from our first-order attitudes to subject them to critical scrutiny, but we rarely take this process further, scrutinizing our second-order attitudes in turn. What matters, fundamentally, is that normative thought is a reflective process, in which we step back from our subjective attitudes and engage in reflection on them; this reflective character is what confers on higher-order attitudes their standing in situations of critical assessment. Higher-order attitudes function as standards of normative assessment, in other words, not because of their nature as desires, but because of the reflective procedures that lead to their formation.

This approach works, however, only in cases in which agents have actually undergone a course of reflection about their first-order desires. Prior to such reflection, the approach suggests, there are no standards for the critical assessment of our motivating

1. See, for example, Harry G. Frankfurt, "Freedom of the Will and the Concept of a Person," in his *The Importance of What We Care About* (Cambridge University Press, 1988), pp. 11–25. [Wallace's note.]

2. Since the higher-order attitudes that are expressed in normative discourse are themselves desires, this version of expressivism can explain the capacity of moral thought to engage the will. But it can also explain why normative thought sometimes fails to give rise to corresponding motivations, since the first-order desires the agent reflects on might be stronger than the second-order desires that normative language expresses. [Wallace's note.]

attitudes, and this is an awkward result. Suppose that in thinking about the question of whether to hang onto the wallet or turn it in at the lost-and-found office, you reach the conclusion that personal financial advantage is not a good reason to keep property that is not rightfully yours. In arriving at this conclusion, you will probably think that you are making a moral and normative discovery, about something that was true all along. It is not that your arriving at this conclusion somehow makes it the case that it is wrong to keep other people's property when it falls into your hands; rather, it was wrong even before you started thinking about the question. But how can the subjectivist make sense of this aspect of moral thought?

One possibility is to appeal to the agent's dispositions. What matters to the normative standing of a given first-order attitude, we might say, is not that the agent has actually endorsed or rejected it through critical reflection, but that the agent is disposed to endorse or reject it through such reflection (i.e., that she *would* endorse or reject the desire if she were to engage in critical reflection on it). Building on this idea, some philosophers have proposed a different way of developing the subjectivist approach, which we might call *dispositionalism*. The dispositionalist holds that normative discourse functions not to express our higher-order attitudes, but to make claims *about* the higher-order attitudes we would arrive at through rational reflection. To say that a lying promise is wrong, on this approach, is to say that one would desire that one not give in to the temptation to make a false promise, if one were to reflect rationally on the matter. When we affirm a normative claim of this kind, we might be expressing our practical attitudes, but we aren't *merely* doing that; we might also be making true statements about a normative subject matter.[3]

Dispositionalism seems to be an advance on expressivism in at least one important respect. It allows us to say that there are normative facts that moral discourse makes claims about, facts that are capable of being discovered when we engage in normative reflection. Moreover, it does this without violating the naturalistic metaphysical commitments of subjectivism. Thus, the normative facts that dispositionalism posits do not involve any weird nonnatural properties, of the kind that would be difficult to locate in the world that the natural sciences describe. Instead, they are facts about the attitudes of human agents, in particular facts about the dispositions of those agents to form higher-order attitudes through critical reflection. Your act of betraying a secret is wrong, on this approach, just in case the following conditional statement is true about you: that you would want yourself not to act on the temptation to betray the secret, if you were to reflect fully on the matter. Dispositional facts of this kind define standards for the normative assessment of the agent's practical attitudes, but the standards are in the relevant sense subjective; they are a matter, fundamentally, of the dispositions of the agents whom the standards regulate.

It is an open question, however, whether dispositionalism can really dispense with normative standards that are independent of the person whose attitudes are subject to assessment. To see this, let's go back to the motivational side of moral thought.

3. See, for example, Michael Smith, *The Moral Problem* (Blackwell, 1994). [Wallace's note.]

Suppose I have arrived at the conclusion that I would want myself to refrain from deception, if I were rational. The dispositionalist says: this judgment *just is* the moral judgment that it would be wrong for me to lie. As we have emphasized, however, moral judgments of this kind are supposed to provide standards not merely for the criticism of our practical attitudes, but for their control; they are practical not just in their subject matter, but in their effect, giving rise to new motivations. But how are judgments about our dispositions to desire things supposed to have this practical effect? The dispositionalist puts motivation into the content of moral judgments, construing them as claims about what we would desire if we were rational. One could form a judgment with this content, however, without having the desire that the judgment is about; how can the dispositionalist bridge this gap?

The natural answer is to appeal to rationality to do this job. That is, dispositionalists often propose the following principle of rationality (or some variant of it):

> It is irrational to judge that I would want myself to do X if I were rational, but to fail to have a desire to do X.

Applying this principle to the case at hand, we get that it would be irrational to judge that it would be wrong to tell a lie for personal advantage, but to fail to desire to act accordingly. Our responsiveness to this standard of rationality, the dispositionalist might then say, is what enables normative reflection to generate new desires. This suggestion is plausible, furthermore, because cases in which we fail to have desires that accord with our normative judgments seem to be paradigm cases of irrationality. If you really believe that you shouldn't lie to your teacher to get an extension on the paper, but you end up doing so anyway, then you are going astray by your own lights; what could be more irrational than that?[4] The problem, for the dispositionalist, is to explain where this principle of rationality comes from. It looks to be a substantive normative standard, one that is prior to and independent of the attitudes that are up for assessment. The postulation of normative standards that are in this way objective, however, violates the subjectivist's most basic metaphysical commitments.

Moral thought involves the application of rational standards, standards that are normative for the agent, in the sense that they properly regulate the agent's critical reflections. The challenge for the subjectivist, as we have now seen, is to make sense of this aspect of moral thought, without recourse to standards that are completely independent of the attitudes of the agents whose reflections they govern. *Constructivism* in moral philosophy can be understood as a response to this problem. On the constructivist view, practical attitudes are subject to scrutiny by reference to critical standards. But those standards are not independent of the attitudes to which they apply; rather, they appropriately govern the subject's deliberation precisely because the subject is already committed to complying with them.[5]

4. Thus, weakness of will is generally understood to be the most flagrant form of irrationality in action. [Wallace's note.]

5. An example of this kind of constructivist view is Christine M. Korsgaard, *The Sources of Normativity* (Cambridge University Press, 1996). [Wallace's note.]

Consider the instrumental principle, which tells us to choose the means that are necessary to achieve our ends. If you intend to go to medical school, for instance, and "Introduction to Organic Chemistry" is a prerequisite for admission to medical school, then the instrumental principle says that you should take the class; your intentions are subject to criticism if you fail to act in this way. But this is because your intention to go to medical school *already* involves a commitment to take the means that are necessary for the attainment of that end. Indeed, the intention to realize the end just is (in part) a commitment to take the necessary means, and hence to comply with the instrumental principle.[6] Constructivists generalize from this example, holding that all of the standards that govern our practical reflections are likewise standards that we are committed to complying with, in virtue of practical attitudes that we have already adopted.

The constructivist approach can be thought of as combining elements of expressivism and dispositionalism. It shares with the former an emphasis on the essential involvement of practical attitudes in the processes of normative and moral reflection. Such reflection takes as its starting point the intentions and desires that we already have, and it attempts to adjust and to refine them through critical thought. We go astray, on this approach, when we fail to live up to our own commitments—as the signers of the Declaration of Independence arguably did, for example, when they condoned practices of slavery while endorsing the principle that "all men are created equal."

Normative reflection can accordingly be understood as a process of figuring out what our commitments really entail, a process that can lead to normative *discoveries*, of the kind the dispositionalist was concerned to make room for. It might take some time for people to come to see that their own commitments (say, about human equality) have the consequence that some of their other attitudes and practices should be rejected or revised. The normative standards that govern the process of critical reflection are thus not restricted to standards whose consequences the agent already explicitly acknowledges. At the same time, the fact that those standards are anchored in the agent's own commitments sheds light on the practical effects of moral thought. For it is in the nature of commitments that they involve an orientation of the will, which moves us to act once we become clear about what the commitments really entail. People who are genuinely committed to the fundamental equality of all people will be moved to abandon and even to fight against practices such as slavery, once they finally face up to the fact that those practices cannot be reconciled with their own moral principles.

This approach represents a promising way of understanding the critical dimension of practical thought, if we accept the subjectivist idea that normative standards are never prior to and independent of the agents whose attitudes they regulate. But

6. Thus, if you realize that "Introduction to Organic Chemistry" is necessary to get into medical school, but you have resolved never to take the course, then it seems you have effectively abandoned your original intention to become a doctor. [Wallace's note.]

the resulting position shares with other forms of subjectivism some consequences that are difficult to accept. Most basically, the constructivist approach makes morality itself hostage to the commitments of the agents whose actions and attitudes are up for assessment. Whether or not it is wrong for me to break a promise or to keep the wallet I have found is ultimately a question of whether, like the Founding Fathers in the case of slavery, I am already committed to moral principles that would prohibit conduct of these kinds.

Kantians in ethics often accept this framework for thinking about moral standards, affirming a generalized constructivism about normativity. They contend that the most basic moral requirements—the moral law or the "categorical imperative"[7]—are universal principles of willing, insofar as they are ones that every agent is *necessarily* committed to complying with. If this claim could be defended, then morality would turn out to represent a set of universal normative constraints on rational agents. But the Kantian claim is exceptionally ambitious, and it has proven very difficult to give a clear and compelling account of the idea that rational agency involves an essential, built-in commitment to follow the moral law. If the Kantian is correct, then it ought to be possible to identify the concrete commitments that villains and scoundrels are betraying when they pursue their reprehensible ends. But does this seem plausible to you? (What is it in the attitudes of the fraudster or the terrorist—people like Bernie Madoff or Timothy McVeigh, say—that would commit them to the basic moral standards that they flout in their actual behavior?)

Those who wish to make sense of morality as a set of nonnegotiable critical standards may therefore need to question the subjectivist framework within which constructivism operates. Perhaps our practical attitudes are answerable to standards that are more robustly independent of the subjects of those attitudes. Before we can accept this objectivist approach, however, we will need to come up with convincing responses to the arguments from metaphysics and motivation canvassed above. Can we make sense of the idea that reality includes irreducibly normative facts and truths, about, for instance, the wrongness of deceptive promises or the impermissibility of exploitation and fraud? How can reflection about such facts and truths reliably give rise to new motivations to action, in the way that we have seen to be characteristic of practical deliberation? These questions continue to push some philosophers back to subjectivist ways of understanding morality, despite the serious difficulties that subjectivism faces in accounting for the critical dimension of normative thought.

7. The categorical imperative is Kant's candidate for the supreme principle of morality, the abstract principle from which our more specific moral duties can be derived. For different formulations of this principle, and Kant's argument that it represents a universal principle of rational willing, see Immanuel Kant, *Groundwork of the Metaphysics of Morals*, ed. and trans. by Mary Gregor (Cambridge University Press, 1997). [Wallace's note.]