

Methods, Goals, and Data in Moral Theorizing

John Bengson
Terence Cuneo
Russ Shafer-Landau

Forthcoming in the *Routledge Handbook of Moral Epistemology* (eds. M. Timmons and A. Zimmerman)

Philosophical methods play a crucial role in philosophical inquiry. When it comes to questions about the nature, status, and content of morality—the special purview of moral philosophy—we look to philosophical methods to help guide the construction of normative and metaethical theories, and to provide the basis for evaluating their individual and comparative merits. One of the tasks of moral epistemology is to determine how this is to be done well.

Here we investigate the construction and evaluation of theories in metaethics, focusing on the nature of the *methods* that should govern metaethical theorizing and their relation to such theorizing.¹ As we'll see, doing so requires attending to both the possible *goals* of metaethical inquiry and (what we'll call) the metaethical *data*—the source-material utilized by good methods to achieve those goals. The main claims about methods, goals, and data for which we'll argue are these:

First, candidate methods for metaethical theorizing must be assessed in light of the epistemic goal(s) of metaethical inquiry.

Second, while there are a variety of epistemic goals that different methods may properly aspire to achieve, several prominent methods face significant challenges when assessed in light of the attractive goal of understanding.

Third, while there are difficult questions about the nature, status, and collection of metaethical data, there are a range of data that must be accounted for by competing metaethical theories.

Fourth, these data possess four basic features, which set them apart from other types of considerations and indicate how and why they serve as the lifeblood of theoretical inquiry.

Fifth, these data should be conceived not as dialectically effective starting points, constitutive features of morality, claims about how morality seems, or descriptions about how moral language is commonly used; rather, they are what metaethicists have epistemic reason to take to be genuine features of morality itself.

Sixth, by utilizing reflection on ordinary moral experience, we are able to reveal what some of these data are.

1. The Tripartite Structure of Metaethical Theorizing

A helpful way to understand the nature of metaethical theorizing is to identify its constituent elements and to reveal the underlying structural relations among them. There are three such elements. There are *inputs* to metaethical theorizing; there are its *outputs*; and there is

¹ Few philosophers have devoted extensive and explicit attention to fundamental questions about theoretical inquiry in moral philosophy, especially metaethics. Treatments of methodology in contemporary moral philosophy have tended to focus primarily on normative ethics and been largely restricted to the particular method that Rawls (1971, 19-20) dubbed “reflective equilibrium,” which we discuss in §1.2. While we'll focus on metaethics, the main elements of our discussion apply *mutatis mutandis* to normative ethics (and many other areas of philosophy).

the *method*, or procedure, that takes one from the former to the latter.² This abstract structure can be realized in numerous ways, depending on the candidate inputs, methods, and outputs that one opts for.

1.1 Outputs

The outputs of metaethical theorizing are theories or views (we'll use these terms interchangeably). These theories possess a variety of properties, including epistemic ones. For example, theories can be coherent, true, or justified; they can afford knowledge, or provide understanding. Theories that possess such positive epistemic properties are, under favorable conditions, the contents of a variety of epistemically significant mental states or attitudes, such as true belief, knowledge, or understanding. These epistemic achievements are among the possible *goals* of theorizing.

It is important to distinguish the outputs of metaethical theorizing from its goals. Outputs are theories. Goals, by contrast, are the agential states we seek to attain through the activity of theoretical inquiry. Standardly, the goals of theorizing are epistemic achievements, such as those mentioned just above. Philosophers have typically failed to be clear about which of these (or other) goals they are pursuing when constructing and evaluating theories. But the differences between these goals are significant, given the dual purpose that such goals serve: they provide standards for assessing the merits of both outputs (theories) and methods. Outputs can be judged by reference to whether they (for example) enable true belief, afford knowledge, or provide understanding. If, for instance, the appropriate goal of metaethical theorizing is understanding, then a theory can rightly be judged inadequate if it doesn't provide that good to an agent who fully grasps the theory. The goals of theorizing will also determine the adequacy of methods, since this depends in large part on whether their successful implementation yields appropriate epistemic achievements. For example, a method might yield true belief (and nothing more). But such a method will fail of its aim if what we are seeking is knowledge, which is more demanding than true belief.

Some goals are more appropriate to one activity rather than other. When the activity is theoretical inquiry, undertaken for its own sake, it's plausible to restrict the range of *proper* goals to genuine epistemic achievements. However, it can't be assumed that there is a unique proper goal for theoretical inquiry; there may be many such goals. It follows that there might be no such thing as *the* uniquely correct method of metaethical theorizing, as different methods might be calibrated to achieve different proper goals.

That said, some such goals are especially important, namely, those whose realization is sufficient to successfully resolve theoretical inquiry—reaching the point at which, to put it simply, there is no more work to be done. We call these *ultimate* proper goals, of which there may perhaps be several. Not all proper goals are ultimate. For example, true belief is not an ultimate goal of theoretical inquiry, since such belief can be unwarranted and still leave open many of the most significant questions about the domain. In fact, even if we attained a set of beliefs that were not simply true, but were also coherent, justified, or even all three of these at once, one could hardly deem inquiry complete. There would remain more work to do, for such a set does not by itself guarantee the sort of *theoretical understanding* required to successfully resolve inquiry.

² This is the second stage of theoretical inquiry (theorizing); the first stage (data collection) consists in collecting the inputs to the second stage. We'll speak of a *method* for theorizing and a *procedure* for collecting the inputs to such a method. We discuss the details of both stages below.

We elucidate some central features of such understanding below, but for now, think of it as the sort of epistemic achievement that provides comprehensive and systematic illumination of its target. One important reason to regard theoretical understanding as an ultimate proper goal of theorizing is that stronger epistemic achievements, such as absolute certainty, are unnecessary to successfully resolve theoretical inquiry. After all, even were we to fall short with respect to such achievements, that would not by itself impugn the success of our inquiry, so long as we had achieved genuine understanding with respect to the central questions in the domain under investigation.³

To summarize, the outputs of method are theories that possess various epistemic properties. The goals of such methods are mental states or attitudes that have such theories as their contents. When all goes well, a method will resolve theoretical inquiry in a fully satisfactory way, realizing an ultimate proper goal. Among the ultimate proper goals is theoretical understanding. In the next section, we consider what method is and sketch what it takes for a method to achieve an ultimate proper goal such as understanding.

1.2 Methods

We understand a method to be a set of instructions, or criteria, for theory construction and evaluation. A method will take some set of inputs and, when successfully applied, will deliver an output that realizes one or more goals. Some goals ‘correspond’ to a method, in one of two senses: either the method was designed to facilitate their realization, or there is something else about the method itself that makes those goals especially fitting. Call these the method’s *corresponding goals*. We’ll say that a method is *valid* just in case satisfying its criteria (or following its instructions) realizes its corresponding goals; a method is *proper* just in case satisfying its criteria realizes a proper goal of theoretical inquiry; and a method is *ultimately proper* just in case satisfying its criteria realizes an ultimate proper goal of such inquiry. Naturally, what is wanted is a method that has all three of these properties: we’ll call such a method *sound*.

As these remarks indicate, the bare notion of method is to some extent a black box, whose contents are its instructions or criteria. Since our objective here is not simply to present the abstract notion of method, but also to identify options for metaethical theorizing, and to show how to evaluate various candidate methods, we propose to work through a few examples. We’ll focus on austere versions of four methods prominent in metaethics. Although these methods are non-exclusive, each is meant to be complete, in that nothing more than following its instructions is required to achieve its corresponding goal. A brief examination of these methods will allow us to pinpoint some of the challenges that they face.

Consider, first, what we’ll call the “method of analysis,” which, focusing on clarification, instructs metaethicists to define or analyze central ethical terms, concepts, or properties:⁴

³ Note that the same considerations do not apply to knowledge—in the standard sense of, roughly, non-accidentally true justified belief—since it is possible to know various truths in a domain without having the sort of illumination constitutive of understanding. This means that in the absence of understanding, the presence of knowledge would leave theoretical inquiry unresolved; if this is correct, then knowledge is not an ultimate proper goal. Independent support for our assessment that understanding is an ultimate proper goal is provided by the defenses of understanding given by Elgin (1996, 122ff.), Kvanvig (2003, ch. 8), and Pritchard (2010), among others. Our assessment is compatible with understanding being an elevated type of knowledge (e.g., higher-order systematic knowledge-why); while this isn’t our view (Bengson forthcoming), we needn’t oppose it here.

⁴ A seminal instance of this method in action is Moore’s *Principia Ethica* (1903, ch. 1). See also Ewing (1947) and Hare (1954). For more recent versions, see Jackson and Pettit (1995), Jackson (1998, ch. 5), and Finlay (2014). While the

Method of Analysis: When constructing a theory, a theorist ought to articulate and justify particular analyses of all of the domain’s central terms, concepts, or properties that meet some sufficiently high standard (e.g., being necessarily coextensive, being intensionally correct, being theoretically serviceable). The best theory is the one whose proposed analyses meet this standard to the highest degree relative to rivals.

A second method, focusing on justification by argument, is what David Chalmers has called the “method of argument,” which calls for theorists to promote their view by putting together the rationale for it:⁵

Method of Argument: When constructing a theory, a theorist ought to assemble an adequate rationale, by formulating arguments for the theory’s answers to the central questions about the domain, as well as arguments that respond to relevant challenges, where the premises and inferences of these arguments meet some sufficiently high standard (e.g., being certain or self-evident, being scientifically or logically well-confirmed, being shared by members of an ideal audience subsequent to extended critical examination). The best theory is the one whose central theses are the conclusions of arguments whose premises and inferences meet this standard to the highest degree relative to rivals.

A third method, focusing on explanation, is what we’ll call the “method of parsimony”:

Method of Parsimony: When constructing a theory, a theorist ought to identify a set of propositions about the domain that realize, to the greatest extent possible, simplicity and explanatory scope (i.e., explanations of everything that must be accounted for). The best theory is the one that achieves the greatest extent and balance of simplicity and explanatory scope relative to rivals.⁶

A fourth method, focusing on systematization, is what John Rawls called “wide reflective equilibrium,” which has been a mainstay of discussions of philosophical methodology, and is

method of analysis takes a variety of forms—some proponents privileging ordinary usage, others formal machinery from contemporary linguistic theory, still others the broadly functional style of analysis known as Ramsification—our characterization in the text abstracts away from these details in order to make explicit its core commitments. Similarly for the other methods characterized next. Similarly for our formulations of the other methods below. This explains why we have described them as ‘austere.’

⁵ Chalmers (2014, 16); *cp.* van Inwagen (2006, Lecture 3). There are many possible—and quite diverse—examples of adherence to this method in metaethical theorizing; candidates include Gewirth (1978), Brandt (1979), Mackie (1977), Nagel (1986), Korsgaard (1996), Shafer-Landau (2003), Huemer (2005), Street (2008), Cuneo (2007), and Wedgwood (2007). A version of the method of argument invoking a comparative, cost-benefit standard is pursued by Enoch (2011, §1.4); *cp.* Schroeder (2007, §11.2).

⁶ This method is highly influential in contemporary metaphysics. For an example in metaethics, see Gibbard (2003, xii), who maintains that his antirealist “hypothesis explains the phenomena—and no normative realism that extends beyond the hypothesis is needed.” *Cp.* Harman (1977), Blackburn (1993, pt. II), Joyce (2001, 168), and Olson (2014, esp. 147-8).

often cited enthusiastically by moral philosophers wishing to clarify their methodological commitments.⁷

Method of Reflective Equilibrium: When constructing a theory, a theorist ought to achieve coherence between various particular judgments (e.g., considered judgments regarding specific cases) and beliefs in general principles (e.g., universally quantified propositions) that address all of the central questions about the domain, through a reflective process of modification, addition, and abandonment of either the particular judgments or principles in case of conflict (with each other, or with any of one's other relevant convictions). The best theory is the one that achieves such coherence to the highest degree relative to rivals.

It might be that the method of analysis yields knowledge of definitions, the method of argument yields justified beliefs, the method of parsimony yields beliefs with a high probability given the evidence, and reflective equilibrium yields a coherent set of judgments. Still, there are live questions about whether any of these methods is valid or proper.⁸ Furthermore, if none of these methods yields a theory that provides understanding or any other ultimate proper goal (recall, e.g., our earlier observation that true, justified, or coherent belief, and even their conjunction, may fail to qualify as such), then none of these methods is sound.

One way to support our doubts about the soundness of these methods is to consider what a method would have to be like in order to yield, as output, a theory that provides a great deal of understanding of a given domain. Such a method would have to yield a theory that possesses at least four characteristics. First, the theory must possess a high degree of *accuracy*, since inaccurate theories will yield only *misunderstanding*. Second, the theory must be *reason-based*, in the sense that it is positively supported by considerations that favor its truth. Third, the theory must be *robust*, answering a multitude of questions about the most important features of the domain under investigation. Fourth, the theory must be *orderly*, not simply offering such feature-specific answers but also affording a broader view of the domain and how it hangs together—for example, by exposing systematic connections among those (and other) features.⁹

We have no particularly good reason to think that any of the austere methods considered above delivers theories that have all four of these characteristics. *Prima facie*, nothing in the method of analysis or the method of argument ensures its outputs will be robust or orderly; the method of parsimony does not guarantee outputs that are simultaneously reason-based, robust, and orderly; and the method of reflective equilibrium (infamously) fails to promise accuracy or guarantee outputs that are reason-based in the indicated sense.¹⁰ Notably, not all proponents would resist our conclusion that these methods fail to secure understanding. For instance, Rawls himself described the method of reflective equilibrium as merely uncovering the doctrine that is

⁷ See, e.g., Sayre-McCord (1996), Depaul (1998), and Scanlon (2002, 149). *Cp.* Timmons and Horgan (2006).

⁸ For example, Bonevac (2004) provides reason to doubt that the method of reflective equilibrium is valid. Many standard criticisms challenge the method's propriety; see, e.g., Hare (1973), Brandt (1979, 19-21 and 1990), Raz (1982), Copp (1985), Stich (1988), Cummins (1998), Kelly and McGrath (2011), and McPherson (2015).

⁹ We identify two further characteristics in Bengson, Cuneo, and Shafer-Landau (in progress-a).

¹⁰ Some proponents of these methods have attempted to address related worries, but we do not find these efforts compelling, partly for the reasons given in Bengson, Cuneo, and Shafer-Landau (in progress-a, §3 and in progress-b, §3).

“most reasonable for us” to accept; similarly, David Brink emphasizes that this method provides justification, even though it may not facilitate stronger epistemic achievements, which would (in our terminology) render it unsound.¹¹

Do our concerns here imply that there is no sound method for metaethical theorizing? We do not draw this conclusion, but acknowledge that work on this topic remains to be done.¹²

1.3 Inputs

We now turn to the inputs of method. Such inputs can be usefully regarded as the *data* of theorizing; they are the materials that a method must account for when issuing its outputs. Such data possess four basic features, which will play important roles in the ensuing discussion.

First, they are *starting points* for theoretical reflection on a domain, bearing an asymmetrical structural relation to subsequent theorizing. Specifically, data are inputs to, not outputs of, such theorizing, the latter generating theories in light of the former, and not vice versa.

Second, data are *inquiry-constraining* with respect to a domain, functioning to anchor a given theoretical inquiry to its subject matter—what a given type of theorizing is about (i.e., what it purports to provide coherent, true, or justified beliefs, knowledge, or understanding of). By saying that they ‘anchor’ inquiry to its subject matter, we intend to convey two points. First, the data operate as the basic means by which theorists access that subject matter. Second, investigations of a domain that entirely ignore the data are likely to be off-track with respect to that domain and, consequently, to fail to achieve any ultimate proper goals.

Third, data are *collected*. While not itself a controversial claim, there is controversy regarding the collection of data, especially over appropriate procedures with which to pursue such collection. Among the candidate procedures for data collection are those that utilize such sources as intuition, introspection, common sense, ordinary experience, induction from experience, linguistic judgments, and observations in controlled scientific experiments.¹³ Of course, there is room to be more or less restrictive on this matter; some philosophers recognize only one of these as part of an appropriate procedure for data collection, while others embrace an indiscriminate pluralism that allows for all such procedures, no one of which is judged to be invariably better than another. A moderate option is a discriminate pluralism that recognizes multiple procedures possessing varying degrees of authority. One reason to favor this moderate position in metaethics is that metaethical data will probably contain both empirical and a priori elements (hence, pluralism), and it is plausible that some procedures will be epistemically better than others (hence, discriminateness).

Fourth, data are *neutral*, in the dual sense that they function as *common currency* among theorists, while also being *fallible* to the extent that a particular datum might be mistaken. Because this feature of data has come under attack by challenges to the very idea of data (stemming from worries about the theory-ladenness of data, to which we’ll respond in the next

¹¹ See Rawls (1980, 534) and Brink (1989, 140-1).

¹² We ourselves attempt to do that work in Bengson, Cuneo, and Shafer-Landau (in progress-a, §4).

¹³ Gathering data with these sources (intuition, etc.) can be done in myriad ways, for example, through reflection on thought experiments, discussion with friends, examination of historical events, scrutiny of a linguistic corpus, surveys, review of scientific journals, and so forth. There are thus candidate *sources* of data, as well as candidate *techniques for employing* those sources. The conjunction of a source and a technique constitutes what we are calling a ‘procedure’ for data collection.

section), it is worth registering that there are at least three reasons for regarding data as neutral in this dual sense.

In the first place, data must not unduly stack the deck in favor of a particular theory, but must instead be admissible to theorists of diverse persuasions. These theorists may differ about how to precisify the relevant data, which are often vague, and disagree as well about which theory best handles the data. (This is a reason for thinking that data function as common currency.)¹⁴ Second, it is always possible for an input to theorizing—a datum—to be denied that status at a later time. This happens when, for instance, theorists acquire reason to view a specific datum as invalid or erroneous (hence, as a source of ‘error in the data’). Third, it is always legitimate for theorists to question whether the data align with genuine features of a domain, instead of being mere noise, or somehow off-base. (These last two points jointly speak in favor of regarding data as fallible.)¹⁵

We’ve identified four basic features of data, the inputs to all theorizing. But the notion of data is not without its critics. We regard it as a virtue of the preceding characterization that it makes sense of the objections and concerns that have been pressed against this notion. In the next section, we’ll illustrate this point by discussing what is arguably the most prominent criticism of them all.

2. The Theory-Ladenness of Data (and Some Related Phenomena)

Philosophers have long worried that data are not neutral starting points but instead are theory-laden; all inputs to theorizing, in metaethics and elsewhere, are themselves in some ways infused with or partial to one theory or another.¹⁶ If that were so, the worry continues, the structural relation between data and theorizing that we have described would be directly threatened, as would the idea of neutrality.

There are several things to say in reply to this concern. The first begins with the observation that a datum’s being “infused with” a theory and its being “partial to” a theory are importantly different. The latter needn’t itself conflict with our conception of data as neutral starting points, as there are familiar cases of such partiality that present no challenge to data functioning as such among theorists of a domain. That physical data favor relativistic theory over Newtonian theory, for example, does not imply that the data unduly stack the deck in favor of relativistic theory, or that the data are the results of—rather than inputs to—such theorizing. Likewise, “infusion” needn’t be problematic, provided that the infusion originates in

¹⁴ Our notion of common currency is distinct from, and does not imply, the type of consensus to which Williamson (2007, 209ff.) objects when discussing a thesis he labels “Evidence Neutrality,” in the context of a forceful critique (with which we are in broad agreement; see §3.3 below) of the tendency to “psychologize” data. The thesis of “Evidence Neutrality” invokes a notion of neutrality on which φ is neutral only if φ is “in principle uncontentionally recognizable as such,” so that—according to Williamson—you have good evidence only if you are “able to persuade all comers, however strange their views, that you have such good evidence” (210-14). By contrast, our notion of common currency comes with no such entailment. Consequently, our notion may help to preserve the role of a data set as “a neutral arbiter between rival theories” (210), a role that Williamson himself applauds.

¹⁵ By viewing data as fallible, we do not rule out that there is a distinct, infallibilist use of the term ‘data’ on which the sentence ‘data cannot be mistaken’ is true. We do not oppose this and other alternative uses of the term. The point we wish to emphasize is that any understanding of *inputs to method* must allow that a given input may be mistaken (whether or not it is called a ‘datum’). Even those who indicate a strong preference for an infallibilist use of ‘data’ and related terms standardly acknowledge this possibility (see, e.g., Williamson 2007, 209-10).

¹⁶ See, e.g., Hanson (1965, ch. 1).

commitments from well-supported theories in domains beyond the one under investigation.¹⁷ No doubt this sometimes occurs; after all, data are collected, and a good source of data collection may utilize such theories in order to do its job well. In short, that data sometimes favor one theory over another, or incorporate various sorts of theoretical commitments, is perfectly compatible with all four features of data enumerated above.

While that point is sufficient to answer worries about theory-ladenness, we also note a second reply. Suppose, as seems plausible, that some data are partial to certain metaethical theories over others. For instance, suppose the intimate connection between moral judgment and action favors certain antirealist views over realist ones, while the possibility of having true moral beliefs and gaining moral knowledge favors success theories over error theories. Realists should and do acknowledge the former datum; error theorists should and do concede the second. But so long as the method that works with the data is itself neutral among competing metaethical theories, in the sense that it allows for the rejection of various data under certain conditions, then all is above board.

And that is because, third, such a method should not regard the data it works with as sacrosanct, in need of preservation come what may. As indicated above, the data are fallible starting points for theoretical efforts. Those efforts may determine, when all is said and done, that some of the data are mere appearances that do not survive critical scrutiny. This is the verdict issued for instance by error theorists with regard to the datum that we have true moral beliefs and some moral knowledge. Alternatively, theorists may accept the data, while rejecting the contention that these data favor a given theory over another. Many moral realists opt for this route after reflecting on the datum that posits a close connection between moral judgment and action.

Setting aside the question of theory-laden data, we should also acknowledge the prospect of what might be called the ‘method-ladenness of data.’ In general, a method will constrain the data (or what qualifies as such) to a certain extent. To illustrate, consider the method of reflective equilibrium. According to this method, the inputs, or data, are myriad particular judgments and beliefs in general principles (and not, say, a type of qualia or desire). The fact that a method—this one or any other—constrains data by identifying criteria for data selection needn’t imply anything fishy. In particular, it does not imply that a proper method will shape the nature, content, or scope of the data it acknowledges as inputs.

By contrast, the data have a profound influence on what qualifies as a sound method. That data and method are connected in this way is not illicit but an innocent consequence of what we term the ‘data-sensitivity of method.’ For example, given that data are inquiry-constraining, any method that allows inquirers to entirely ignore the data couldn’t qualify as sound. (Below we describe another example, when discussing our favored conception of data.)

Above we identified four basic features of data; we’ve now noted how both theory and method may legitimately interact with data, and vice versa. The foregoing, however, leaves open some controversial issues regarding how best to conceive of data and their role in theoretical inquiry. We devote the next section to considering some of these issues, including several that deeply shape the character of metaethical theorizing.

¹⁷ *Cp.* Boyd (1988, 206-7). Another possibility is that it is a *precisification* of the data that receives the infusion, in which case the data themselves retain their status as neutral starting points.

3. Four Conceptions of Data

There are at least four very different conceptions of the data that philosophers have adopted. These conceptions offer informative characterizations of data, proposing conditions under which something qualifies as a datum. Each can be evaluated in light of the extent to which it preserves the four basic features of data identified above (in §1.3). Examining these conceptions will allow us not only to explore candidate views of the nature and status of data, but also to sharpen our understanding of how data are related to the subject matter of metaethics.

We begin with two conceptions of data that we believe to be inadequate for straightforward reasons. We then examine a third conception that avoids those problems, though it faces others. Finally, we explain our preferred conception—which privileges the epistemic status of data—and note a few of its principal virtues.

3.1 Two faulty conceptions: *Dialectical and metaphysical*

According to what we call the *dialectical conception*, φ is a datum if and only if, and because, φ is a claim (or what is expressed by a claim) that inquirers (considered collectively) provisionally agree upon as central to the domain in question.¹⁸ This conception has the virtue of being well-positioned to make sense of the neutrality of data; it also meshes with the observations that data are starting points and collected. But it does not square with the idea that data are inquiry-constraining, for it does not guarantee that investigations of a domain that entirely ignore the data (as the dialectical conception thinks of them) are likely to be off-track with respect to that domain.¹⁹

In contrast to the dialectical conception, the *metaphysical conception* of data holds that φ is a datum if and only if, and because, φ is a (claim regarding a) feature that is genuinely constitutive of a domain; the data give the domain both its existence and character.²⁰ This conception, unlike the former, preserves the idea that data are inquiry-constraining, for it ensures that investigations of a domain that entirely ignore the data are likely to be off-track. The metaphysical conception is also consistent with the fact that data are collected. However, it is not obviously compatible with the idea that data are starting points—specifying the features that are constitutive of a domain is often a result of inquiry, rather than one of its inputs. Perhaps more flagrantly, the metaphysical conception implies that the data are never mistaken, which directly contradicts the fallibility—hence, the neutrality—of data.

¹⁸ See, e.g., Heney (2016, 26). *Cp.* the “pragmatic view” discussed by Williamson (2007, 238). (Heney also conceives of her view as a pragmatic one.) The expression ‘considered collectively’ is ours; it has two functions. First, it brackets the commitments that individual inquirers incur due to personal endorsement of particular theories of the domain. Second, it respects the epistemic position enjoyed by (members of) a collective towards some claim even in the face of a subset of the collective whose members personally lack that epistemic position with respect to that claim (perhaps because they happen to endorse idiosyncratic views—for example, idealism or eliminativism about the mental—that function for them as defeaters of the reasons possessed by their colleagues).

¹⁹ After all, the dialectical conception allows that data (to borrow a memorable phrase from Richard Rorty) closely track “what our peers let us get away with.” With perhaps the singular exception of inquiry regarding what our peers let us get away with, this falls far short of anchoring inquiry to its subject matter in the relevant sense.

²⁰ For two possible examples in metaethics, see Korsgaard (1996, Lecture 1) and Joyce (2001, ch. 2), who hold that certain data regarding features of reasons—their excellence or authority—are constitutive of morality. *Cp.* Gibbard (1990, 32-3), who treats the intimate relation between moral judgment and action similarly. We are not confident whether these or any other metaethicists embrace the metaphysical conception for *all* data. But, importantly, our objections to this conception hold even when the conception is restricted to a subset of data.

3.2 The psycho-linguistic conception

More popular than the previous two conceptions is what we will call the *psycho-linguistic conception* of data, which holds that φ is a datum if and only if, and because, φ belongs to a domain-specific class of psychological or linguistic claims (or whatever such claims express). In the paradigm cases, such data are (or are expressed by) statements about how things seem or how language is aptly or commonly used. For example, within metaethics, it is widely accepted that moral values often *seem* to be objective and that it is apt to *say*—perhaps when discussing interrogation techniques—“It’s a fact that torture is wrong.”

The psycho-linguistic conception holds that *all* data are like this.²¹ To illustrate, consider a more precisely formulated version of a putative metaethical datum alluded to above, namely:

Practicality: Moral judgments have marks of practical attitudes: for example, they guide and motivate action.

This formulation makes no explicit reference to how things seem or how language is used; it tells us about a putative fact regarding the intimate connection between moral judgments and action. Now consider:

Practicality _{ϕ} : It *seems* that moral judgments have marks of practical attitudes: for example, they *seem* to guide and motivate action.

Practicality _{λ} : It is apt to *say* (directly or by implication) “Moral judgments have marks of practical attitudes: for example, they guide and motivate action.”

These claims make explicit reference to how things seem or how language is used; they do not concern a putative fact regarding the connection between moral judgments and action. The psycho-linguistic conception understands data to take the form not of Practicality but Practicality _{ϕ} or Practicality _{λ} .

The principal motivation for the psycho-linguistic conception’s restriction of the data to psychological or linguistic claims is to satisfy neutrality. If the data consist in claims about how things seem or how language is used, then the data can function as common currency among otherwise rival metaethical theorists, whose disputes about the nature and status of morality do not impede their agreement about how morality seems or how moral language is used. And if the data consist in claims about how things seem or how language is used, then the data may be fallible, failing to depict genuine features of the metaethical domain, which are not entirely psychological or linguistic. This virtue of the psycho-linguistic conception is, however, intimately intertwined with what is perhaps its greatest vice, which can be stated as a dilemma focusing on the relation between the metaethical data and the subject matter of metaethics.

Suppose, first, that the psycho-linguistic conception holds that the data themselves are the subject matter of metaethics. In that case, the subject matter of metaethics would consist entirely

²¹ Some metaethicists hold restricted versions of the psycho-linguistic conception, which allow that some data are not psycho-linguistic, even though most are (see, e.g., Finlay 2014, 121). The dilemma we will advance below applies equally to restricted versions, so long as they leave a sizeable gap between the metaethical data and the subject matter of metaethics.

in how things seem or how language is used. But that is seriously mistaken, for it leaves out huge swaths of the metaethical domain, concerning (for example) the actual nature and status of moral concepts, propositions, properties, and facts, including the relations between such things and their non-moral counterparts; the actual nature and status of moral reasons, including their relations to moral values and whether they are ever categorical; the true nature of moral judgments, including their connection with action; the possibility of justified moral belief and moral knowledge; and so on.

In fact, all prominent metaethical theories have recognized in practice that the subject matter of metaethics is not exhausted by how things seem and how language is used. Expressivist theories, for instance, offer accounts not simply of Practicality_ψ or Practicality_λ but also of the connection specified in Practicality. And when criticizing rival cognitivist views, they charge not that these rivals are unable to account for Practicality_ψ or Practicality_λ, but that they cannot account for the connection specified in Practicality (or cannot do so as well as expressivist theories do). Expressivists thereby engage in apt theorizing, which takes its object to be not mere psychological or linguistic claims of the sort recognized by the psycho-linguistic conception, but to include features, or putative features, of morality itself.

This brings us to the second horn of the dilemma. Suppose that the psycho-linguistic conception were to *deny* that the data are the subject matter of metaethics, holding instead that the data bear some other relation to what it is that metaethical theorizing is about. This ensures that the data are neutral, and thus sustains the main motivation for the psycho-linguistic conception. It also sidesteps the objection raised just above, since it avoids the consequence that the subject matter of metaethics consists—implausibly—in how things seem or how language is used. However, we see three problems with this way of proceeding.

First, this approach posits a sizeable gap between the data and the subject matter of metaethics. Given that data are inquiry-constraining, functioning so as to enable metaethical theorizing to access the nature and status of morality, this gap cannot remain impassable; it needs to be bridged, so as to connect metaethical theorizing, whose inputs—on the psycho-linguistic conception—are psycho-linguistic data, to the metaethical domain, which is not entirely psycho-linguistic. Here we find a deep tension within the psycho-linguistic conception. On one hand, it requires there to be a substantial gap between the data and the subject matter of metaethics, both to secure neutrality and to avoid omitting huge swaths of the metaethical domain. On the other, it also needs there *not* to be a substantial gap, in order to ensure that metaethical theorizing can access the subject matter of metaethics.

This brings us to the second point: while there may be ways to resolve this tension (most obviously, by treating the data as an epistemic indicator of the subject matter²²), they risk simply moving the bump in the rug by generating conflict elsewhere. To appreciate this, suppose we were to cross the gap by endorsing a bridge principle that licenses a transition from the relevant psycho-linguistic claims to the (not entirely psycho-linguistic) subject matter of metaethics; such a principle would facilitate access to the latter via the data. Suppose, for example, a theory were to embrace the principle that things generally are as they seem, which would allow theorists to get at (say) the connection in Practicality via Practicality_ψ.²³ This principle is transparently

²² Though such a move risks collapsing into the epistemic conception, discussed below.

²³ We will not dwell on the concern, which this illustrates, that the psycho-linguistic conception's posit of a sizeable gap between the metaethical data and the subject matter of metaethics introduces a gratuitous epicycle to

incompatible with the fundamental commitments of various metaethical theories. No proponent of error theory or expressivism, for instance, could endorse it. To the contrary, these theories are committed to rejecting any such principle, as they are committed to the claim that, when it comes to morality, things are often, and perhaps quite generally, not as they seem. The challenge for proponents of the psycho-linguistic conception, then, is to identify a principle that bridges the gap between psycho-linguistic data and the subject matter of metaethics but isn't in tension with the commitments of their own metaethical theories. It is unclear whether this challenge can be met.

The third point adds to and intensifies the challenge posed by the conjunction of the first two points. As just discussed, in order for theorists to access the subject matter of metaethics via psycho-linguistic data, they will have to commit themselves to a bridge principle of some kind. But any such principle is bound to be contentious. At the very least, it will not be neutral in the sense of functioning as common currency among theorists. The worry, then, is that the psycho-linguistic conception secures neutrality by positing a gap between metaethical data and the subject matter of metaethics, which can be bridged only by introducing further commitments that are not neutral at all. The psycho-linguistic conception gives with one hand what it takes away with the other.

3.3 *The epistemic conception*

We now turn to the *epistemic conception* of data, which holds that φ is a datum if and only if, and because, φ is a claim (or what is expressed by a claim) that inquirers (considered collectively) are in a good epistemic position to take to identify genuine features of a given domain. So, for example, Practicality—not merely Practicality $_{\psi}$ or Practicality $_{\lambda}$ —is a metaethical datum just in case, and because, metaethicists are in a good epistemic position to take it to identify a genuine intimate connection between moral judgments and action. Different versions of the epistemic conception will analyze ‘good epistemic position to take’ in different ways. According to our preferred version, the analysis invokes the familiar category of *reason for belief*, which we understand as a defeasible epistemic status (i.e., it is possible for a reason for belief to be outweighed or extinguished by sufficiently weighty countervailing considerations). Let us call this the *epistemic reason conception* of data.

Before considering a few of the virtues of this conception, let us first observe that it is non-committal in some important respects. It is compatible with different explanations of why inquirers have reason to take a domain to be as the data characterize it as being (e.g., epistemic conservatism, phenomenal conservatism, dogmatism, process reliabilism, proper function theory, virtue theory, safety theory, subjective Bayesianism, etc.). It is also compatible with different views about the strength of the reasons to take a domain to be those ways, and how much is required to defeat them. Further, it is compatible with different views about what is required to legitimately reject—by defeating the reason for—any given datum or set of data.

Unlike some of its rivals, the epistemic reason conception preserves neutrality: the data function as common currency but are also fallible. It also makes sense of the fact that data are inquiry-constraining: it provides inquiry with an *epistemic anchor* (i.e., a reason-based relation) to its subject matter, thereby implying that investigations of a domain that entirely ignore the data

metaethical theorizing—*gratuitous* because it can be avoided compatibly with neutrality (the principal motivation for the psycho-linguistic conception), as shown by our discussion of the epistemic conception below.

are likely—epistemically likely, owing to the reasons theorists possess—to be off-track with respect to that domain.

The epistemic reason conception also makes sense, in interesting and important ways, of the fact that data are collected, by explaining central features of such collection. To appreciate this, consider the felt need to justify the sources we employ in data collection, which flows from the demand to use only those sources that provide *reason* to take the domain to be the ways they tell us that it is. In the paradigm case, when collecting data, inquirers believe (or assume) that their sources of data have positive epistemic status, providing them with reason to take the domain to have certain features. For example, if inquirers think they should collect data from the outputs of our best physical theories, that is because they hold that those outputs are ones they have reason to believe. Or if inquirers think they should collect data via intuitions about thought experiments, that is because they hold that intuitions about thought experiments provide reasons for belief. The epistemic reason conception straightforwardly accounts for this dimension of data collection.

It also explains controversy over candidate sources of data collection. Some would insist, for instance, that to be a source of data, we must have independent warrant for regarding its verdicts about features of the domain to be reliable. Others would reject this. These healthy disagreements make sense only to the extent that a source of data functions to provide reason for believing that the domain in question is a certain way. The controversy arises because the question of what it takes to satisfy this condition is both open and often extremely difficult.

Another virtue of the epistemic reason conception is that it explains the possibility of disputes over the data in a domain. Naturally, theorists may disagree about what they (considered collectively) have epistemic reason to take to be genuine features of the domain under investigation.

The epistemic reason conception has several further important implications for the nature, status, and collection of inputs to theorizing.²⁴ It also bears on the propriety of methods. Given that inquirers have reason to take data to be genuine features of the domain under investigation, any method of theorizing that allows inquirers to reject the data without justifying their rejection couldn't qualify as sound. In effect, the epistemic reason conception explains our earlier observation that the data have a profound influence on what qualifies as a sound method (what we've labeled the data-sensitivity of method, to which we'll return below).

4. The Data for Metaethical Theorizing

In the previous section, we argued that data are what inquirers have (defeasible) reason to take to be genuine features of a domain. Here we identify considerations that satisfy this description in metaethics, thereby buttressing our preferred conception and demonstrating how it applies in practice.

Our point of departure is the following ordinary scenario:

Collegiality: A colleague of yours has been in a bad car accident, sustaining major injuries. To ease the burden on her and her family, you pledged several weeks ago to

²⁴ For instance, it implies that those who restrict the metaethical data to a specific set of considerations, such as ordinary usage, are guilty of chauvinism. For they are tacitly assuming that no other considerations could supply epistemic reason to take the metaethical domain to be one way rather than another. Our view avoids such chauvinism, by allowing that data are epistemically-supported considerations, whatever those happen to be.

provide a meal for them on a particular date. You have forgotten all about this, and the automatic reminder delivered by your calendar this morning—the date has arrived—comes as a surprise. You give the matter some thought. You reckon that your colleague and her family probably won't go hungry tonight if you don't provide a meal. But after running through a number of such scenarios, it strikes you that while there are alibis available, you should cancel the other plans you made yesterday and prepare the meal you committed to providing: given your previous commitment, this is quite clearly what the situation demands. And so you judge that this is what you ought to do. Being moved by this verdict, you set out to prepare the meal.

Assuming that there is nothing especially unique about this scenario, we can extract from it at least four candidates for metaethical data.²⁵

First, we have reason to think that there is the relation between your judgment about what to do and your subsequent action, which indicates the practical character of the former. Your judgment is capable of guiding your decision-making and behavior (in much the way that grasp of a recipe can guide what you do). It also moves you (perhaps in conjunction with certain of your desires or commitments), at least to some extent, to act. This inspires the following datum, which was already mentioned above:

Practicality: Moral judgments have marks of practical attitudes: for example, they guide and motivate action.²⁶

Second, at the same time, we have reason to think that your judgment about what to do also has tell-tale marks of being a descriptive attitude, such as belief. It is a response to how things strike you, in which you find yourself “coming down” on a verdict about a way the world is: the situation is such as to demand a certain type of response on your part. This judgment is a way of categorizing or classifying the world. Like paradigm beliefs, it is also fitted to play various other roles. For example, we have reason to think that its content can enter into further inferences and other sorts of logical constructions (e.g., conditionals such as *if I don't prepare a meal, then I'll have failed to do what I ought to do*). Moreover, this content can be felicitously described as true or false in what seem to be perfectly straightforward uses of these terms (e.g., if queried about your responsibilities, it would not be odd for you to say, “Yes, it's true—I ought to prepare a meal”). This motivates a second datum:

Descriptivity: Moral judgments have marks of descriptive attitudes: they are classificatory, truth-evaluable, and apt for inference.

Third, we have reason to think that your judgment about what to do also has epistemic dimensions, in that it is natural to think that you may occupy a better or worse position for grasping the moral demands that apply to you. For instance, if things go well, your judgment will

²⁵ *Cp.* Horgan and Timmons's (2006, 222-3) list of the “phenomena of morality,” which appear to function as data.

²⁶ This is not intended to be a universal generalization. Nor must it be read as a generic. We use the plural ‘moral judgments’ because it is easy to see that many moral judgments are like the one in Collegiality. A similar point applies to other data.

be justified and may even constitute knowledge and facilitate understanding about what you ought to do. These observations motivate a third datum:

Grasp: Moral agents grasp, or are duly placed to grasp, moral reality.

Fourth, we have reason to think that there is one correct judgment (or one limited set of correct judgments) to make in response to your situation; this implies that you can make a mistake about what morality demands of you. This gives rise to a fourth datum:

Fallibility: Not any judgment in response to a moral question will do; we can make moral mistakes.

There are several points to make about the four candidate data we've identified. Obviously, they do not exhaustively characterize moral thought, discourse, and reality (or even the moral dimensions of Collegiality). Nor are they highly determinate, since they—like nearly all data—are fairly vague and open to interpretation, refinement, clarification, or elucidation; given that data are *inputs* to theorizing, this is as it should be. At the same time, they arguably bear the four marks of data we identified earlier (in §1.3). First, they function as starting points of metaethical theorizing, identifying some central features of ordinary moral life that subsequent theorizing must take into account. Second, they constrain metaethical inquiry, helping to anchor it to its subject matter; indeed, any metaethical theory that entirely ignored these data would be safely deemed off-track. Third, they are collected, issuing from a legitimate source, utilizing reflection on ordinary moral experience. Finally, they are neutral in the dual sense introduced earlier: they are fallible and are well-suited to function (indeed, they *do* function) as common currency among metaethical theories, even if some theories have ultimately rejected one or more of them as mere appearances.

Let us close by observing that, while the four data we have listed are neutral, they are not methodologically innocuous, at least not if we are right about the data-sensitivity of method. If a metaethical theory were to entirely ignore these data, or to reject them without defeating the reason we have to take them to describe genuine features of morality, that theory would be in worse shape relative to its rivals that take these data into account (either by accepting them or by adequately justifying their rejection). It follows that metaethical theories that flout these data in the ways just mentioned would issue from unsound methods. For such methods would have failed to recognize that data are both inquiry-constraining and backed by reasons, making those methods ill-suited to realize the ultimate proper goals of metaethical inquiry.²⁷

References

- Bengson, John. Forthcoming. "The Unity of Understanding." In Stephen Grimm, ed. *Making Sense of the World*. Oxford University Press.
- Bengson, John, Terence Cuneo, and Russ Shafer-Landau. In progress-a. "Method in the Service of Progress."

²⁷ Many thanks to Lauren Davidson, Mark Timmons, Aaron Zimmerman, and participants in a seminar at Harvard for helpful comments on earlier versions of this material.

- Bengson, John, Terence Cuneo, and Russ Shafer-Landau. In progress-b. "Methodological Trends in Contemporary Metaethics."
- Blackburn, Simon. 1993. *Essays in Quasi-Realism*. Oxford University Press.
- Bonevac, Daniel. 2004. "Reflection Without Equilibrium." *Journal of Philosophy*, 101: 363-88.
- Boyd, Richard. 1988. "How to be a Moral Realist." In Geoffrey Sayre-McCord, ed. *Essays on Moral Realism*. Ithaca: Cornell University Press: 181-228.
- Brandt, Richard. 1979. *A Theory of the Good and the Right*. Oxford University Press.
- Brandt, Richard. 1990. "The Science of Man and Wide Reflective Equilibrium." *Ethics*, 100: 259-78.
- Brink, David. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge University Press.
- Chalmers, David. 2014. "Why Isn't There More Progress in Philosophy?" *Philosophy*, 90: 3-31. Reprinted in Ted Honderich, ed. 2015. *Philosophers of Our Times*. Oxford University Press: 347-70.
- Copp, David. 1985. "Morality, Reason, and Management Science: The Rationale of Cost-Benefit Analysis." *Social Philosophy and Policy*, 2: 128-51.
- Cummins, Robert. 1998. "Reflection on Reflective Equilibrium." In Michael Depaul and William Ramsey, eds. *Rethinking Intuition*. Rowman & Littlefield Publishers, Inc.: 113-28.
- Cuneo, Terence. 2007. *The Normative Web*. Oxford University Press.
- Depaul, Michael. 1998. "Why Bother with Reflective Equilibrium?" In Michael Depaul and William Ramsey, eds. *Rethinking Intuition*. Rowman & Littlefield Publishers, Inc.: 293-309.
- Elgin, Catherine. 1996. *Considered Judgment*. Princeton University Press.
- Enoch, David. 2011. *Taking Morality Seriously*. Oxford University Press.
- Ewing, A.C. 1947. *The Definition of Good*. Macmillan.
- Finlay, Stephen. 2014. *Confusion of Tongues. A Theory of Normative Language*. Oxford University Press.
- Gewirth, Alan. 1978. *Reason and Morality*. University of Chicago Press.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Oxford University Press.
- Gibbard, Allan. 2003. *Thinking How to Live*. Harvard University Press.
- Hanson, N.R. 1965. *Patterns of Discovery*. Cambridge University Press.
- Hare, Richard M. 1954. *The Language of Morals*. Oxford University Press.
- Hare, Richard M. 1973. "Rawls' Theory of Justice." *Philosophical Quarterly*, 23: 144-55; 241-51.
- Harman, Gilbert. 1977. *The Nature of Morality*. Oxford University Press.
- Heney, Diana B. 2016. *Towards a Pragmatist Metaethics*. Routledge.
- Horgan, Terry and Mark Timmons. 2006. "Morality Without Moral Facts." In James Dreier, ed. *Contemporary Debates in Moral Theory*. Blackwell: 220-38.
- Huemer, Michael. 2005. *Ethical Intuitionism*. Palgrave.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Clarendon Press.
- Jackson, Frank, and Philip Pettit. 1995. "Moral Functionalism and Moral Motivation." *The Philosophical Quarterly*, 45: 20-40.
- Joyce, R. 2001. *The Myth of Morality*. Cambridge University Press.
- Kelly, Thomas and Sarah McGrath. 2010. "Is Reflective Equilibrium Enough?" *Philosophical Perspectives*, 24: 325-59.
- Korsgaard, Christine. 1996. *The Sources of Normativity*. Harvard University Press.
- Kvanvig, Jonathan. 2003. *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press.
- Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong*. Penguin Books.
- McPherson, Tristram. 2015. "The Methodological Irrelevance of Reflective Equilibrium." In

- Christopher Daly, ed. *The Palgrave Handbook of Philosophical Methods*. Palgrave MacMillan: 652-74.
- Moore, G. E. 1903. *Principia Ethica*. Cambridge University Press.
- Nagel, Thomas. 1986. *The View from Nowhere*. Oxford University Press.
- Olson J. 2014. *Moral Error Theory*. Oxford University Press.
- Pritchard, Duncan. 2010. "Knowledge and Understanding." In *The Nature and Value of Knowledge: Three Investigations*, co-authored with Alan Millar and Adrian Haddock. Oxford University Press: 3-88.
- Rawls, John. 1971. *A Theory of Justice*. Harvard University Press.
- Rawls, John. 1980. "Kantian Constructivism in Moral Theory." *Journal of Philosophy*, 77: 515-72.
- Raz, Joseph. 1982. "The Claims of Reflective Equilibrium." *Inquiry*, 25: 307-30.
- Sayre-McCord, Geoffrey. 1996. "Coherentist Epistemology and Moral Theory." In Walter Sinnott-Armstrong and Mark Timmons, eds. *Moral Knowledge*. Oxford University Press: 137-89.
- Scanlon, Thomas. 2002. "Rawls on Justification." In Samuel Freeman, ed. *The Cambridge Companion to Rawls*. Cambridge University Press: 139-67.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defense*. Oxford University Press.
- Schroeder, M. 2007. *Slaves of the Passions*. Oxford University Press.
- Stich, Stephen. 1988. "Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity." *Synthese*, 74: 391-413.
- Street, Sharon. 2008. "Constructivism about Reasons." In Russ Shafer-Landau, ed. *Oxford Studies in Metaethics*, volume 3. Oxford University Press: 207-46.
- van Inwagen, Peter. 2006. *The Problem of Evil*. Oxford University Press.
- Wedgwood, Ralph. 2007. *The Nature of Normativity*. Oxford University Press.
- Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Blackwell.