

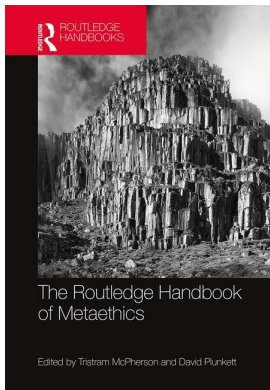
This article was downloaded by: 10.3.97.143

On: 01 Apr 2023

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



The Routledge Handbook of Metaethics

Tristram McPherson, David Plunkett

Naturalistic Realism in Metaethics

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9781315213217.ch2>

Peter Railton

Published online on: 28 Aug 2017

How to cite :- Peter Railton. 28 Aug 2017, *Naturalistic Realism in Metaethics from: The Routledge Handbook of Metaethics* Routledge

Accessed on: 01 Apr 2023

<https://www.routledgehandbooks.com/doi/10.4324/9781315213217.ch2>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Naturalistic Realism in Metaethics

Peter Railton

How do ethical thought and practice relate to our experience and understanding of the world more generally? This question has many dimensions. Metaphysically, what is the place of value in a world of facts? Conceptually, how are claims about what *ought* to be the case related to claims about what *is* the case? Epistemically, how can our ethical beliefs be justified, when they often seem to be based upon “intuitions” that cannot be adjudicated by empirical experimentation or logical proof? And semantically, how can our thoughts and words successfully connect with putative ethical properties or facts? Such challenges are not limited to metaethics and can be found in some form right across the meta-normative realm, but it is with respect to ethics that attempts to meet them have been most developed, so that will be our principal focus here.

“Naturalistic realism” names a family of metaethical approaches that seek to provide a unified response to these challenges by grounding ethics in the natural world in a way that leaves no “unexplained residue”—whether metaphysical, conceptual, epistemic, or semantic—that could be considered indispensable to the well-foundedness of ethical distinctions (e.g., between intrinsically good and bad lives, morally right or wrong actions, and just or unjust practices) or to their practical significance (e.g., as objective values or rational standards to guide deliberation and action).

This would include no unexplained *super-natural* or *non-natural* residue—no need to appeal to divine commands, punishment in an afterlife, or a realm of absolute Ideas. Plato’s argument in the *Euthyphro* (ca. 399–395 BCE/2002) has convinced most philosophers that, even if divine beings exist, still, ethics has an independent standing. But Plato did not think that ethics could do without the “Forms”—abstract, ideal, non-natural Ideas that stand behind our concepts. And many contemporary non-naturalists agree with Plato at least that far—the ethical realm is “just too different” from the ordinary world of experience in its principles, ways of knowing, or modes of thought and language (though whether this difference should be interpreted metaphysically is a matter of debate among non-naturalists; see Moore, 1903; Gibbard, 2003; Enoch, 2011; Parfit, 2011).

However, at least since Plato's pupil Aristotle (ca. 350 BCE/1999), naturalistic realists have countered that we can find all we need to vindicate the claims of ethics without going beyond the natural world, which of course includes humans as natural beings. What is required for vindication? At the heart of the issue are two characteristics of ethics that appear to be in some tension with one another: *factuality* and *normativity*. On the one hand, judgments about the quality of a life, the rightness of an action, or the fairness of a practice share many features with prosaic, paradigmatically factual judgments about the natural world. Ethical and prosaically factual claims are spoken of as true or false, have a shared propositional logic, and combine seamlessly in thought and language. Speakers making both kinds of judgments appear to presuppose that they are talking of a common subject matter, such that their disagreements can be genuine disputes, and such that they can be called upon to justify themselves in terms of reasons that are shareable and not wholly dependent upon their particular standpoint or opinion. Ethical and prosaically factual judgments alike characteristically go beyond immediate observation and invoke properties and relations that are general and not directly observable—such as causal connections, laws of nature, intrinsic value, and reasons for action. Further, in both domains, these properties and relations have a *modal* character and thus support expectations about the future and reasoning about hypothetical and counterfactual situations. Finally, in both domains, we strive for impartiality and objectivity, seeking to avoid personal or social bias and refusing to accept brute appeals to authority. In all these respects, judgments with ethical versus prosaically factual or naturalistic content are as alike as peas in a pod.

On the other hand, ethical judgments appear to have a *normative character* or *force* that prosaically factual judgments do not—they *prescribe* or *commend* rather than merely describe. So it seems anomalous at best, and incoherent at worst, to assert in all sincerity that an action is morally wrong or unjust while being indifferent to whether it is performed (philosophers differ considerably, however, on how to understand normative character or force, or its relation to ethical judgment; see David Faraci and Tristram McPherson's chapter "Ethical Judgment and Motivation").

A naturalistic vindication of ethics would thus need to accommodate both its factuality and its normativity. Historically, philosophers have developed several ways of accomplishing this. Aristotelian science saw nature in terms of essences and causes—including "final causes" that explain the behavior of objects and creatures in terms of a teleological striving to realize their distinguishing essential characteristics. In sentient creatures, this striving takes an *affective, intrinsically motivating* form, such that animals experience pleasure or happiness in activity that is in accord with their essence. Dogs, for example, find pleasure in the full development of their distinctive canine nature—eating, running, hunting, sociality, and nurturing their young. Humans, by contrast, find pleasure in the full development of their distinctively rational, political nature. Human virtues, including self-regulation by individual and shared deliberation, are grounded in the facts of human nature and at the same time intrinsically motivating—an inseparable part of achieving happiness (*eudaimonia*) (Aristotle, ca. 350 BCE, 1999: 1098–102). Naturalistic realism inspired by this Aristotelian linkage of essence with flourishing, and of human flourishing, in turn, with virtue and rational deliberation, can be found today in contemporary neo-Aristotelianism (Foot, 1977; McDowell, 1979).

A second tradition in naturalistic realism, and the one that will be the principal focus of this entry, got its start when Thomas Hobbes applied the Galilean Revolution to

human science, banning Aristotelian “final causes” and using an empirical “resolutive-compositive” method that traces complex dynamical phenomena to the interactions of simpler constituent elements. For Hobbes, humans pursuing the fulfillment of their desires will, if rational, accept and abide by reciprocal constraints akin to morality as their best hope for peace and happiness (Hobbes, 1651/1994). David Hume improved upon Hobbes’ psychology, adding both a theory of sentiments, which made humans sensitive to a wider range of values, and an inherent capacity for general sympathy, which meant that humans could adopt, and be moved by, perspectives other than their own—including the perspective of the general interest (Hume, 1738/1978). Combining these two features meant that people were disposed to approve of traits of character beneficial to others as well as the self (“natural virtues”), and to be responsive in feeling and action to mutually beneficial norms of justice and property (“artificial virtues”). And John Stuart Mill (1863/2001) gave greater theoretical and empirical content to these broad Humean ideas by introducing informed preferences as a standard for assessing personal well-being (utility), and the maximization of well-being overall as a standard for assessing social norms.

Naturalistic realism in Anglophone ethics went into decline in the early twentieth century, in part due to the influence of G.E. Moore (1903), which will be discussed below, and in part due to the rise of positivism in philosophy and the sciences, which dismissed Aristotelian essences and Millian utility alike as unverifiable and thus empirically meaningless. Those seeking a naturalistic explanation of ethics thus tended toward “emotivism” (Ayer, 1936) or “expressivism” (Gibbard, 2003) rather than realism, or, more radically, toward an “error theory” of ethics (Mackie, 1977). But the rejection of positivist strictures in the natural and social sciences in the second half of the twentieth century, along with a series of parallel developments in metaphysics and the philosophy of language, made possible the contemporary revival of naturalistic realism. Let us, then, begin with a brief discussion of realism, and then of naturalism.

REALISM

What is it to be a *realist* about a given domain of thought or practice? Like other philosophical terms of art, the meaning of “realism” is controversial and involves an interplay between common-sense language and the demands of conceptual clarity and theory construction. In contemporary philosophy, realism is often formulated in linguistic terms—realists about the external world may formulate their view by saying that our ordinary language of objects successfully refers—but it is important to see that they typically seek to defend the existence of a world of objects that is quite independent of how we speak about it.

We will speak here of realism *about a domain*, since what realism requires varies by domain. Realism about the external world is sometimes put in terms of its *mind-independent existence*, but this will hardly do for realism about mental states—or, according to most naturalists, for realism about ethics. Rather than seek a defining characteristic of realism, we will consider a branching series of *yes-no* choices, such that *yes* answers lead to successively greater degrees of realism about the domain in question.

First, realists about a domain tend to hold that (1) statements in that domain are capable of truth or falsity (or, more generically, of correctness or incorrectness) and that (2) at

least some affirmative existential statements in that domain are in fact true (or correct). Early *emotivists* about ethics denied (1), and more recently ethical *error theorists* have accepted (1) but denied (2). For comparison, an atheist or error-theorist about religion might accept that (1) the statement “God exists” is capable of truth or falsity, but deny that (2) this statement is true. Of course, statements are only true or false when interpreted, and so the next branching question is whether (1) and (2) hold when (3) statements in that domain are given a *literal* interpretation. For example, a *non-literal* theist might claim that “God exists” is true, but when asked for the meaning of this statement say, “It means that I face the future with confidence.” Such a position satisfies (1) and (2), but not (3). A non-literalist about ethics might claim that the statement “Torture is wrong” means “I hereby resolve to shun anyone who tortures or who tolerates torturing,” or that talk of such a statement being true is “a useful fiction” for promoting social coordination. If someone accepts that affirmative ethical claims are true when literally interpreted, the next question is whether she accepts that (4) truth in this context is *opinion-independent*. For example, a theist who claims that “‘God exists’ is true for anyone who sincerely believes it” is denying opinion-independence, and similarly for a theist who claims that “Whether God exists is relative to a culture.” Finally, there is the question whether (5) these literal, opinion-independent truths can provide a *substantive* explanation of our beliefs or practices in this domain. A stage-(5) realist about the external world might say, “The external world exists apart from whatever we believe, but nonetheless can shape our experience via sensation—thus our veridical perceptual beliefs can be said to be true *because* the world is the way it is.” Call this a *tracking* account. By contrast, someone who accepts (1)–(4) but rejects (5) might give a *projectivist* account (cf. Blackburn, 1998): “Truth is not a substantive relation between our minds and the world that can be invoked to do explanatory work—to say that ‘*p*’ is true is to say no more than *p*, and thus to explain what it is for *p* to be true is simply to explain what it is to *think* or *assert* that *p*.” Tracking realists start off with an account of what it is for *p* to be true, and then seek to illuminate the perceptual or cognitive processes that permit the reality of *p* to be part of a causal explanation of why *p* is believed, while projectivist realists (who may call themselves *quasi-realists*) seek to show how an account of what it is to believe that *p* obviates the need for such an explanation and the metaphysics that would support it (see Dreier, 2004; McPherson, 2015; see also Terence Cuneo’s chapter “Quasi-realism”).

With this understanding of the landscape of realism, let us consider naturalism and its varieties.

NATURALISM: TWO BASIC ORIENTATIONS

Whatever their differences, philosophical naturalists tend to share a sense that the empirical, theoretical, and mathematical sciences afford our best-developed and most successful examples of inquiry aimed at obtaining a fundamental understanding of the world and ourselves. Moreover, the sciences have developed powerfully predictive explanations of many of the phenomena traditionally dealt with speculatively in philosophy, and in the process overturned long-held philosophical convictions concerning such questions as the nature of space and time, the ultimate constitution of matter, and the source of life and “design” in living organisms. Naturalists urge that we learn from this history

that philosophical inquiry should be pursued in tandem with empirical science, not as a synthetic *a priori* “first science.” This is not to say that philosophy can make no distinctive contributions to science—in the last half-century, e.g., Bayesian epistemology and decision theory have influenced a range of disciplines, and theories of speech acts and conversational norms have played an important role in linguistics. On the contrary, it is to say that philosophers should seek to develop theories that can be contributory parts of the larger explanatory enterprise of science.

Naturalists differ in approach. *Substantive naturalists* start off with a stand on ontological questions, according to which, strictly speaking, all bona fide fundamental entities, properties, relations, or facts are, or are reducible to or grounded in, entities, properties, relations, or facts of the kind posited in the natural sciences. (For short, we will speak of “properties and facts.”) For further discussion, see Gideon Rosen’s chapter “Metaphysical Relations in Metaethics”). By contrast, *methodological naturalists* are committed in the first instance to a *project* rather than a metaphysics. Insofar as possible, they seek to work compatibly with the methods of the natural and social sciences as they tackle fundamental philosophical questions. Whether this will, in the end, result in the conclusion that the “natural world” is all the reality there is, is for them a question to which we have no *a priori* answer. (For additional discussion, see Daniel Nolan’s chapter “Methodological Naturalism in Metaethics.”)

NATURALISM AFTER MOORE: ANALYTIC AND NON-ANALYTIC APPROACHES

In the founding text of twentieth-century metaethics, *Principia Ethica*, G.E. Moore attacked what he called the “naturalistic fallacy” of attempting to define ethical concepts in naturalistic terms (Moore, 1903). (He also claimed to find a similar fallacy in attempts to define ethical concepts in super-natural or theistic terms.) His “open question” argument is meant to show why such definitions must fail. Take any alleged naturalistic definition of an ethical concept, e.g., of the moral goodness of an action in terms of whether, relative to alternatives, it maximizes net happiness overall. Moore argues that the philosopher advancing this account takes herself to be making an informative statement with which someone could disagree while displaying an adequate grasp of the concept MORALLY GOOD. (Moore framed his argument in terms of *properties* rather than concepts, but, for reasons to be discussed below, a formulation in terms of concepts seems to offer a more convincing version of the argument.) Moore concluded that we have a *sui generis* concept of “morally good,” distinct from any proposed naturalistic characterization of what *makes* acts or outcomes good.

Moore was *not* saying that we cannot give definite, even naturalistic, answers to the question of what makes acts or outcomes good or right—indeed, he proposed such answers himself (1903, 1968). Rather, he claimed that any such answers must be *synthetic* rather than analytic—they involve substantive claims (what he would call “intuitions”) about how we ought to live, and no one should be allowed to foist such claims upon us in the guise of “definitions.” This he took to refute previous naturalisms in ethics, such as the utilitarianism of Mill (1863/2001).

Moore’s argument has been enormously influential, though formulating it clearly and assessing its significance have proven difficult, and remain controversial (see Nicholas

Laskowski and Stephen Finlay's chapter "Conceptual Analysis in Metaethics"). Contemporary naturalistic replies fall into two categories: some agree that developing and defending their accounts must go beyond conceptual analysis, while others reject Moore's argument as based upon a mistake and thus propose a new form of analytic naturalism.

Non-analytic naturalists can draw upon several strategies. Some ethical concepts for which we have no naturalistic equivalent or "reduction" might be able to figure *in their own right* in legitimate scientific explanations. For example, interpersonal cooperation in a group or society can break down for various reasons—lack of resources, conflicting interests, external incursion, etc. But in many cases, the most illuminating explanation of breakdown will be that the existing arrangement was *unfair* to one or more of the parties—not "thought to be unfair," but *genuinely* unfair, even if this was not recognized at the time. UNFAIRNESS, then, might be a worthwhile organizing concept in diagnosing social dynamics, and so earn its way into the vocabulary of social-scientific explanation (for other examples, see Sturgeon, 1995). This may be especially plausible in the case of so-called *thick* normative concepts, in which descriptive and normative elements each figure essentially (see Debbie Roberts' chapter "Thick Concepts").

More generally, non-analytic naturalists can argue that even if some ethical *concepts* are not reducible to natural concepts, it may nonetheless be true that natural *properties* and *facts* suffice to explain the truth of claims involving these concepts. Consider the origin of the word "concept," which lies in two Indo-European roots, *con* + *kept*, "with + grasp/take." This suggests an image of concepts as "handles" with which we mentally take hold of the objects of which we wish to speak. Concepts can achieve this in a variety of ways, corresponding to different "modes of presentation" of their objects. Thus, the same individual can be "taken hold of" under the definite-descriptive concept THE TALLEST MAN IN THE BAR, the "baptized" concept JASON BEMIS, or the compound indexical concept THAT MAN TO YOUR RIGHT—three ways of conceiving this individual, none reducible to another, but the man *himself* is a single human being situated in the causal order, not a description, a name, or an indexical. If I am a bar-keep wondering how to reach a bottle on a high shelf, then the fact that this man satisfies the definite-description concept THE TALLEST MAN IN THE BAR presents him in a way that is directly relevant to my concern. But if I need to identify him to keep track of his tab, regardless of who else comes into the bar, then the individual concept JASON BEMIS is more apt. Importantly, the same individual can be also be picked out by *normative* concepts, which have special relevance to other purposes, for example, THE LEAST TRUSTWORTHY MAN IN THE ROOM or A MORAL AGENT.

Concepts are "modes of mental presentation," but *properties* are features of the world, e.g., dimensions of connection or similarity that unite a class of objects. Properties include substance types (*being iron*, *being made of proteins*), causal-explanatory or functional types (*being an acid*), structural types (*possessing bilateral symmetry*), dynamic types (*having angular momentum*), genetic-historical types (*being a parent of twins*, *being a Homo sapiens*), and also evaluative or normative types (*being intrinsically desirable*), and so on. Unlike concepts, the man in question actually *is* (or *is not*) made of proteins, bilaterally symmetric, a parent of twins, a *Homo sapiens*, etc.

So, while concepts have roles in thought, properties have roles in the world. The ordinary-language concept ACID and the microphysical concept PROTON DONOR pick out the same underlying property with the same causal-explanatory powers, even though

the concepts are not analytically equivalent. The existence of this shared, underlying property tells us that the ordinary-language concept ACID nonetheless picks out something real and explanatory, which accounts for why the ordinary-language concept ACID is so useful, and how it can be used (e.g., in “Lemon juice tarnishes metal because it is acidic”) to convey genuine explanatory information.

A non-analytic naturalist about ethics can likewise seek to determine whether ordinary-language concepts like MORALLY GOOD or MORALLY RIGHT might pick out underlying real and explanatory properties, which would enable us to see why these concepts are useful and important in organizing the experience of life. Hume, for example, suggests that the property of *being morally good*, as attributed to traits of character, is constituted by *being such as to contribute reliably to the general interest* (1738/1978: 580). This property–constitution relationship could explain why grouping traits under the concept MORALLY GOOD or VIRTUOUS picks out features of a kind of fundamental importance in social groups, and thus similar concepts are likely to be found across many cultures. This “origin” of the moral distinction between good and bad traits of character cannot be discovered simply by analyzing the ordinary-language concept MORALLY GOOD or VIRTUOUS—it requires a theory of “human nature” and social dynamics, such as Hume develops in the *Treatise*. Arguably, it is this sort of explanatory theory, rather than an analytic definition, that Aristotle, Hobbes, Hume, Adam Smith (1759/1994), Jeremy Bentham (1780/2007), Mill, and other important naturalistic realists had in mind in the first place—they were trying to give an informative, vindicatory account of *what it is to be* morally good (or morally right, etc.) and *why this has an important action-guiding role* in human life, rather than an analysis of ordinary-language ethical concepts. Their accounts are “definitions” only in the sense of the “real” or “reforming” definitions typical of scientific theories, which seek to account for everyday features of the world in terms of underlying entities, structures, and processes—e.g., explaining macroscopic *heat* in terms of *molecular kinetic energy* (for realisms of this kind, see Railton, 1986a, 1986b, 1989; Boyd, 1988; Brink, 1989). Such accounts are *synthetic* rather than analytic and thus are not affected by Moore’s open question argument.

Some contemporary naturalists take issue with the “open question” argument itself, arguing that it involves a question-begging assumption that analytic equivalences must be *obvious* to competent speakers of a language. “A sister is a female sibling” may seem, on a moment’s reflection, obvious. But what about “A circle is a continuous set of points in a plane equidistant from a single point in that plane”? Given the complexity of ethical questions, it is likely that concepts such as MORALLY GOOD or MORALLY RIGHT will be yet more complex than CIRCLE, so we should not assume that spelling out their definitions will be obvious. Proponents of the “Canberra Plan” in metaethics (Jackson, 1998) argue that ethical terms can be defined in terms of the roles they play in mature ethical theory and practice, in much the same way that Frank Ramsey (1931) and David Lewis (1970) argued that theoretical terms in science can be defined via *their* roles in mature scientific theory and practice. By using quantification over properties, it is possible to formulate a statement of a well-developed theory that preserves the inferential relations and connections with observation or paradigm cases of the theory, but which does not contain any problematic theoretical terms (“unobservables” in science, or normative terms in ethics). MORAL GOODNESS, for example, could be defined as picking out whatever property most closely satisfies the role occupied by “moral goodness” in our

best-developed ethical theory, including common-sense truisms about moral goodness (e.g., that it is impartial) and paradigm cases (e.g., that the relief of suffering is morally good, other things being equal). The resulting role-based definition can be thought of as giving the complex a “job description” to be satisfied by anything that would count as moral goodness. It might not be immediately obvious that whatever property best satisfies this job description *is* moral goodness, they argue, but if one sees this equation as giving the correct analysis of our concept of MORAL GOODNESS, then one can evade the Moorean critique of analytic naturalism—it won’t be an “open question” whether a (possibly complex) natural property satisfying the job description is morally good. (For a non-analytic version of the “job description” strategy, see Railton, 1993.)

SUPERVENIENCE AND EXPLANATION

Let us return to our opening worry: Why think that it should be possible *at all* to give a naturalistic account of the ethical, given the categorical difference between *is* and *ought*? A key part of the answer lies in the profound ways in which the ethical properties are tied to the non-ethical, and to natural facts and properties in particular.

Imagine two individuals identical in all natural features, situated in two worlds also identical in their natural characteristics. Philosophers across the meta-normative spectrum largely agree that this naturalistic specification already assures that these individuals will have the same *evidence* and *practical reasons*—if one individual’s beliefs are justified by his experience to date, then so are the other’s; if one has good reason to take a certain action in his current situation, then so does the other. We cannot replicate natural features and then freely add or subtract normative features. This relationship between the normative and the natural is usually called *supervenience*. Very roughly, a feature or group of features *S* supervenes upon another feature or group of features *T* if, necessarily, fixing *T* also fixes *S* (see Pekka Väyrynen’s chapter “The Supervenience Challenge to Non-Naturalism,” for discussion of the many forms and strengths of supervenience).

A familiar example of supervenience is the *average value* of a variable in a population. For instance, once the individual ages of all the members of a given sample have been fixed, then so is the average age. The average of a variable in a population belongs to a class of higher-order statistics—including the median, the mode, the variation, etc.—which carry important information about a population in their own right, but which someone would fail to understand if he thought of them as *sui generis* or self-standing features that could be altered without changing any of the first-order features in the sample.

Indeed, these statistical relationships are actually “stronger” than mere supervenience, since they are asymmetric—the first-order distribution of ages fixes the average, median, mode, and variation in the sample, but not vice versa. (By contrast, *having a triangular shape* supervenes symmetrically on *having a trilateral shape*.) The higher-order value statistical values thus can be said to be *dependent upon* and to *hold in virtue of* the first-order values—the higher-order values are, in Moore’s terminology, “consequential” from the first-order values, a notion of “consequence” that is asymmetric but need not be causal (Moore, 1968).

In these statistical examples, asymmetric dependence is mediated by the analytic definitions of the higher-order variables, but in many cases, it is not. For example, the macroscopic *heat* of a gas supervenes asymmetrically upon the kinetic behavior of the

molecules composing it—a given microstate determines a unique heat value for the gas as a whole, but not the reverse. Thus we say that macroscopic heat holds in virtue of, or is “consequential” from, the gas’s kinetic microstate, even though there is no purely analytic definition connecting the two.

This example illustrates two other important dimensions of asymmetric dependence: the *specificity* of the relation and the *explanatory status* of the dependent property or magnitude. For example, while the macroscopic heat of a given volume of gas is fixed by the *total microstate* of the universe, it also is fixed more specifically by the motion of *its own constituent molecules*—a key discovery of kinetic theory. This discovery does not supplant but rather vindicates the ordinary concept HEAT in many of its applications—e.g., invoking HEAT in an explanation of the ability of steam to melt ice or power a mill—since it shows that these uses of HEAT pick out a genuine underlying magnitude that accounts for these phenomena. Citing macroscopic heat is not simply redundant, however, since the ability to melt water or power a mill will be the same for two volumes of fluid or gas if they have the same *average* molecular kinetic energy, even though their molecular compositions differ. This similarity in ability is a *higher-order* characteristic, and the HEAT-invoking account captures this explanatory *generality* and *unification*, which would be missing from a mere conjunction of the two individual micro-level explanations.

Asymmetric, specific dependency in the absence of an analytic definition is a common relationship between our everyday categories of experience and underlying natural features. This holds for categories with normative dimensions as well. *Health*, for example, can be thought of as a normatively important higher-order property of organisms that depends asymmetrically and specifically upon a constellation of natural features of body and mind, ranging from cardiovascular sufficiency, to the absence of certain disease agents, to the balance of neurotransmitters in the brain, to physical mobility and resilience in the face of stress. When sufficiently many of these features are in place, the organism is healthy, and poor health can be traced to specific kinds of deficits or failures in such features. We do not need to posit health as a *sui generis* or self-subsistent property, and someone would misunderstand the concept HEALTH if he thought he could improve his pet hamster’s health without changing its mental or physical condition. The concept HEALTH picks out real, complex features of individuals and populations, though the specific details will vary from individual to individual or species to species. HEALTH thus can provide general and unified explanations, and promoting health can be a unified project of inquiry and action (e.g., “Health declined when the medical system broke down, resulting in much-reduced productivity” or “Maternal health is a major determinant of infant mortality in humans and animals alike”). The concept HEALTH is what Richard Boyd (1988) termed a “homeostatic cluster concept,” which collects under a single “handle” or “mode of presentation” a range of underlying physical and mental characteristics that work together in a mutually sustaining way—in this case, to underwrite organismic functionality and freedom from suffering (for criticism of this account, see Horgan & Timmons, 1992; for defense, see Merli, 2002; van Roojen, 2006; and Dowell, 2015). This connection with functionality and quality of experience helps explain why the property of being healthy is of enduring interest in human societies, and why the concept HEALTH can be used to orient shared discussion and inquiry, even among those who disagree about some of the underlying components of health—e.g., the current debate over what range of body mass indices is healthy.

Similarly, naturalistic realists characteristically think of ethical properties as “higher-order” features linked to core areas of shared human concern. Like HEALTH, ethical concepts such as GOOD, RIGHT, or FAIR collect together multiple phenomena that deeply affect our lives—e.g., well-being, cooperation, reciprocity, mutual respect, etc.—under a unifying “mode of presentation” with direct relevance for individual and shared deliberation and action-guidance. At the same time, ethical properties are not self-standing or *sui generis*—ethical facts depend asymmetrically and specifically upon constellations of natural, social, and psychological facts, and someone would show a deficient grasp of the concepts of GOODNESS or FAIRNESS, for example, if she thought she could improve a society’s goodness or fairness while leaving in place all these empirical facts about people’s lives. At the same time, ethical concepts can figure in informative, general explanations of various patterns of behavior and outcomes—e.g., “Rates of compliance with the tax system declined as tax policy became progressively less fair” or “After trying her hand at various things, she settled upon life-coaching because she discovered it was a good life for her” (Railton, 1986a, 1986b).

Naturalistic realists are not alone in thinking that ethical features depend asymmetrically and specifically upon natural features. “Like cases must be treated alike” is a common-sense truism in ethics. In this truism, likeness is not understood trivially in terms of “ethically alike,” but substantively, in terms of likeness in features that can be understood independently of ethical judgment. In practice, these likenesses are typically quite specific—two acts of deception can differ in their location in space and time, the genotypes of the agents, the ambient temperature, and atmospheric pressure, etc., yet still be “like” enough in the intentions, expectations, and harms involved to require the same ethical evaluation or response. Even expressivists, who reject a metaphysical interpretation of supervenience, accept this sort of normative principle of treating like cases alike and thus accept that “fully opinionated” ethical judges are committed to there being some or other (possibly complex) natural features that all acts that are good, or right, share (Gibbard, 2003; for discussion, see Elisabeth Camp’s chapter “Metaethical Expressivism”).

Where naturalistic realism is distinctive is in the *explanation* it affords of this truism, and of the asymmetric, specific dependence of the ethical upon the natural that underwrites it. Just as the concept HEALTH is a normative “mode of presentation” of a constellation of natural features of body and mind that constitute *what it is* to be healthy, our concepts GOOD, RIGHT, and FAIR are normative “modes of presentation” of natural features of lives, acts, and practices that constitute *what is* good, right, or fair. Thus, Hume, for example, saw himself as making the substantive discovery that a combination of enduring human interests and a human capacity for general sympathy explains why we deem some traits virtues and others vices, why we should expect to find such distinctions in all societies, and how such distinctions can be motivating even in the face of self-interest. Moreover, he argues, once we understand this account, we will reflectively approve of our ethical distinctions and practices as corresponding to something real and important. “Thus upon the whole I am hopeful, that nothing is wanting to an accurate proof of this system of ethics” (1738/1978: 618).

To be sure, naturalistic realists need not seek to vindicate *all* aspects of ordinary ethics—this would be impossible given the conflicts in ethical thought found within and across societies. Often, as in the case of Bentham or Mill, naturalistic realists have sought to identify a core basis that explains a great deal of common-sense ethics but also could

support the systematic reform of beliefs and practices—e.g., in the case of slavery or the subjugation of women—that traditionally represent themselves as ethical, but which are in tension with the fundamental characteristics of impartiality, treating like cases alike, and concern to avoid suffering and promote well-being. This works in much the same way that the “real” or “reforming” definitions of mature science seek to identify a core explanatory basis for many of the categories of ordinary experience, but also to eliminate conflations of categories, misleading superficial similarities, and erroneous associations, and to extend our conceptual vocabulary to accommodate novel phenomena.

Such a substantive explanation of supervenience also helps us see how epistemic and semantic access could exist for ethical facts—as we saw in connection with Boyd’s “homeostatic cluster concepts.” Similarly, Hume gives an account of *what it is* for a trait of character to be a “natural virtue” in terms of the trait’s tendency to promote individual well-being and the “general interest” of social groups. This means that our individual and shared life experiences—as well as those of other individuals, other societies, and other times—can provide evidence that enables us to “track” whether a given trait is genuinely a natural virtue. Moreover, we can see how it is possible that more diverse experiences (Mill’s “experiments in living,” 1859/1978), enhanced capacity for empathic simulation, and greater understanding of human psychological and social dynamics can give us greater justification and reliability in these assessments. Social “tracking” accounts of this kind provide an explanation of how our ethical language could succeed in referring to ethical facts, providing a common ground for ethical discussion and disagreement and funding talk of truth and error in ethical judgments.

DIFFICULTIES WITH NATURALISTIC REALISM

It should be obvious from the descriptions given above that naturalistic realism of the “tracking” variety has many hostages to fortune. What follows is a brief survey of some of the difficulties faced by naturalistic realists, along with some preliminary responses.

Naturalistic realism involves a parochial attitude toward ethics. The “tracking” naturalistic realist gives an explanation of supervenience, and of the development of ethical discourse and practice, that draws essentially upon substantive ethical claims. This should not be surprising—we should hardly expect to be able to give a theory of how perception and language put us reliably in touch with features of the world without making some substantive assumptions about what those features are like. Addressing this worry requires that such assumptions be weak, and beg as few questions as possible. Thus, we have assumed that ethics strives for objectivity in the form of impartiality and opinion-independence, that prevention of harm or suffering are among the central ethical concerns, and that cases alike in their natural features must be treated alike by ethical evaluation. This list of features is compatible with all the main contemporary ethical theories, but it does contrast ethics with normative systems of a purely egocentric, ethnocentric, conventional, or divine kind, or of a kind that is entirely indifferent to suffering. Is this too parochial? Perhaps not. Would I count as having successfully taught my children the meaning of FAIR or MORALLY RIGHT if what I have instilled in them is a view according to which these terms apply always and only “because I say so,” and guide action only “because otherwise you’ll be in trouble”?

Interestingly, elementary school children across a range of cultures are already too skilled in making normative distinctions to be taken in by such spurious instruction (for a summary of evidence, see Smetana, 2006 and Turiel, 2006). For example, most children will follow a substitute teacher's instruction to raise both hands before speaking, but balk at his instruction to poke their seat-mates with a pencil to signal the desire to talk. Children explain their resistance in terms of avoiding harm to others, and they exhibit intrinsic motivation to accord such authority-independent norms special seriousness and priority in practice—even intervening to enforce these constraints in third-party interactions when this is at some cost to themselves. Moreover, they distinguish unprovoked harm from harm in the form of punishment for previous acts of harming others. It seems unlikely that all these children will have been given explicit instruction in making these distinctions, or in their relative importance in practice—especially since children can use these distinctions to oppose a parent's or teacher's instructions. Rather, it would appear that they have acquired—perhaps with assistance from innate predispositions (Hamlin, 2013)—an implicit, non-parochial understanding of the distinctive character and importance of the dimensions of normative assessment we have characterized as ethical.

Taking evolutionary theory seriously is incompatible with realism about ethics. Critics have argued on evolutionary grounds that a naturalistic approach to ethics will not sustain realism. While it might be clear why natural selection would have favored the emergence of perceptual systems capable of mapping the local natural environment accurately, why would it have favored the emergence of capacities to detect and be moved by putative objective moral truths? Such truths, we have claimed, would be impartial, abstract, unobservable, and modal. How could our perceptual systems, evolved to operate in the realm of the perspectival, concrete, observable, and actual, provide reliable feedback about such moral truths? And how could it contribute to one's inclusive reproductive fitness to constrain one's thought and action by an impartial concern with the well-being of all affected? (For some examples of evolutionary critiques of realism about ethics, see Joyce, 2001 and Street, 2006.)

We have seen, however, that a naturalistic realist's substantive account of ethical facts can explain how ordinary experience could permit access to the ethical features of lives, traits of character, or actions. To be sure, the moral truths thus learned are impartial, abstract, unobservable, and modal, but so are the causal truths of common sense and science—and so is the distinction between conventional versus authority-independent constraints that elementary schoolers implicitly master. Moreover, we now have very good behavioral and neuroscience evidence for the chief Humean mechanism for learning such moral truths: "general sympathy"—the ability to vicariously experience how things are from others' points of view, with attendant immediate effects on feeling and motivation (Decety & Ickes, 2009). Affective simulation of other points of view—including the points of view of one's future selves—appears to play a key role in human behavior, and individuals deficient in this ability appear to suffer both in personal decision-making and in the formation of mutually beneficial long-term social relations (Blair & Blair, 2009). Affective simulation generates expectations with respect to the course of one's own experience and the likely behavior of others, and these afford a basis for learning by feedback from experience. Infants in their early years seem to map their social as well as physical environment in non-perspectival terms, keeping track of evidence of trustworthiness and helpfulness in the behavior of third parties, and are less likely to accept information or instruction from adults whose third-party behavior has been deceptive or harm-

ful (Doebel & Koenig, 2013; Nguyen et al., 2016; for adult social valuation, see Behrens et al., 2008). In short, humans appear to have evolved to be able to apply and be moved by impartial, abstract, aperspectival, modal, harm-based impartial assessments of behavior (for discussion of evolutionary mechanisms that would explain this, see Nesse, 2007).

Naturalistic realism does not provide a satisfactory account of the normative character of ethical judgments. Expressivists and “projectivist” (or quasi-) realists consider it a decisive advantage for their view that they can establish an *internal* connection between moral judgment and motivation: the state of mind expressed in ethical judgments, they argue, is a motivating one, and not a mere belief purporting to “track” a state of affairs (Gibbard, 2003). But the connection between ethical judgment and motivation is much-contested (for discussion, see Smith, 1994, 1995 and Darwall, 1997; see also David Faraci and Tristram McPherson’s chapter “Ethical Judgment and Motivation”), and, in any event, it is unclear that a connection to *motivation* suffices to capture *normative* force. Consider someone who judges progressive taxation to be just because it equalizes burdens and would be agreed to as a social-insurance scheme by agents who did not know how the “natural lottery” would affect their fates, but who is not personally motivated to pay his share in such a scheme. Is he misusing the word “just”? Would it be a more accurate expression of his state of mind for him to pronounce progressive taxation “unjust”?

The evidence for a *necessary* connection between moral judgment and motivation is not decisive, but surely there should be a *general* connection of this kind, or ethics would have little role in human social existence (see David Faraci and Tristram McPherson’s chapter “Ethical Judgment and Motivation”). The Humean mechanism of “general sympathy” offers a naturalistic explanation of why ethical judgment and motivation would tend to go hand-in-hand: the primary way by which we acquire, retain, refine, or revise ethical beliefs is via mentally simulating actions and outcomes in a way that is inherently sensitive to how those affected are harmed or helped, and that generates unfavorable versus favorable affective responses accordingly. Because this mechanism, when working well, “tracks” genuine good- or right-making features, it affords an explanatory advantage over “projectivist” approaches that lack a general account linking the *substantive content* of ethical judgment with psychic processes capable of yielding ethical beliefs that are *reliable* as well as motivating (Dreier, 2015; see also Joshua Schechter’s chapter “Explanatory Challenges in Metaethics”). Moreover, as noted earlier, Aristotelian and Hobbesian forms of naturalistic ethical realism offer their own accounts of the connection between the content of ethical judgment and motivation.

No account of the normativity of ethical thought and language has emerged that commands wide consensus. What can be asked of naturalistic realism at present is that it exhibit resources for grappling with this difficult question that show some promise. Whether in the end one of these approaches will offer a convincing account that combines factuality with normativity remains to be seen.

REFERENCES

- Aristotle (ca. 350 BCE, 1999) *Nicomachean Ethics*, trans. and ed. by T. Irwin, Indianapolis, IN: Hackett.
 Ayer, A.J. (1936) *Language, Truth, and Logic*, London: Gollancz.

- Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., & Rushworth, M.F.S. (2008) "Associative Learning of Social Value," *Nature* 456: 245–49.
- Bentham, J. (1780/2007) *An Introduction to the Principles of Morals and Legislation*, Mineola, NY: Dover Reprints.
- Blackburn, S. (1998) *Ruling Passions*, Oxford: Clarendon.
- Blair, R.J.R. & Blair, K.S. (2009) "Empathy, Morality, and Social Convention: Evidence from the Study of Psychopathy and Other Psychiatric Disorders," in J. Decety & W. Ickes (eds.), *The Social Neuroscience of Empathy*, Cambridge, MA: MIT Press.
- Boyd, R. (1988) "How to be a Moral Realist," in G. Sayre-McCord (ed.), *Essays in Moral Realism*, Ithaca, NY: Cornell University Press.
- Brink, D. (1989) *Moral Realism and Its Foundations*, Cambridge: Cambridge University Press.
- Darwall, S. (1997) "Reasons, Motives, and the Demands of Morality," in S. Darwall et al. (eds.), *Moral Discourse and Practice: Some Philosophical Approaches*, Oxford: Oxford University Press.
- Decety, J. & Ickes, W. (eds.) (2009) *The Social Neuroscience of Empathy*, Cambridge, MA: MIT Press.
- Doebel, S. & Koenig, M.A. (2013) "Children's Use of Moral Behavior in Selective Trust Discrimination versus Learning," *Developmental Psychology* 49: 462–69.
- Dowell, J. (2015) "The Metaethical Insignificance of Moral Twin Earth," *Oxford Studies in Metaethics* 11: 1–27.
- Dreier, J. (2004) "Meta-Ethics and the Problem of Creeping Minimalism," *Philosophical Perspectives* 18: 23–44.
- Dreier, J. (2015) "Explaining the Quasi-Real," *Oxford Studies in Metaethics* 11: 273–97.
- Enoch, D. (2011) *Taking Morality Seriously: A Defense of Robust Realism*, Oxford: Oxford University Press.
- Foot, P. (1977) *Virtues and Vices*, Oxford: Oxford University Press.
- Gibbard, A. (2003) *Thinking How to Live*, Cambridge, MA: Harvard University Press.
- Hamlin, J.K. (2013) "Moral Judgement and Action in Preverbal Infants and Toddlers," *Current Directions in Psychological Science* 22: 186–193.
- Hobbes, T. (1651/1994) *Leviathan*, ed. by E.M. Curley, Indianapolis, IN: Hackett.
- Horgan, T. & Timmons, M. (1992) "Troubles for New Wave Moral Semantics: The Open Question Argument Revisited," *Philosophical Papers* 21: 153–75.
- Hume, D. (1738/1978) *Treatise of Human Nature*, 2nd edition, ed. by L.A. Selby-Bigge & P.H. Nidditch, Oxford: Clarendon.
- Jackson, F. (1998) *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Clarendon.
- Joyce, R. (2001) *The Myth of Morality*, Cambridge: Cambridge University Press.
- Lewis, D. (1970) "How to Define Theoretical Terms," *Journal of Philosophy* 67: 427–46.
- McDowell, J. (1979) "Virtue and Reason," *Monist* 62: 331–50.
- Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*, London: Penguin.
- McPherson, T. (2015) "What Is at Stake in Debates among Normative Realists?" *Noûs* 49: 123–46.
- Merli, D. (2002) "Return to Moral Twin Earth," *Canadian Journal of Philosophy* 32: 207–40.
- Mill, J.S. (1859/1978) *On Liberty*, ed. by E. Rapaport, Indianapolis, IN: Hackett.
- Mill, J.S. (1863/2001) *Utilitarianism*, 2nd edition, ed. by G. Sher, Indianapolis, IN: Hackett.
- Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press.
- Moore, G.E. (1968) "Replies to my Critics," in P.A. Schilpp, *The Philosophy of G.E. Moore*, Vol. 2, La Salle, IL: Opencourt.
- Nesse, R. (2007) "Runaway Social Selection for Displays of Partner Value and Altruism," *Biological Theory* 2: 143–55.
- Nguyen, S.P., Gordon, C.L., Chevalier, T., & Girgis, H. (2016) "Trust and Doubt: An Examination of Children's Decision to Believe What They Are Told about Food," *Journal of Experimental Child Psychology* 144: 66–83.
- Parfit, D. (2011) *On What Matters*, Vol. 2, ed. by S. Scheffler, Oxford: Oxford University Press.
- Plato (ca. 399–395 BCE, 2002) *Euthyphro*, in G.M.A. Gruber and J.S. Cooper (trans. and ed.), *Plato: Five Dialogues*. 2nd edition. Indianapolis, IN: Hackett.
- Railton, Peter (1986a) "Moral Realism," *Philosophical Review* 95: 163–207.
- Railton, Peter (1986b) "Facts and Values," *Philosophical Topics* 224: 5–31.
- Railton, Peter (1989) "Naturalism and Prescriptivity," *Social Philosophy and Policy* 7: 154–74.
- Railton, Peter (1993) "Noncognitivism about Rationality: Benefits, Costs, and an Alternative," *Philosophical Issues* 4: 36–51.
- Ramsey, F. (1931) "Theories," in *The Foundations of Mathematics*, ed. by R.B. Braithwaite, London: Routledge and Kegan Paul.

- Smetana, J. (2006) "Social-Cognitive Domain Theory," in J. Smetana and M. Killen (eds.), *Handbook of Moral Development*, Mahwah, NJ: Erlbaum: 119–53.
- Smith, A. (1759/1994) *The Theory of Moral Sentiments*, ed. by K. Haakonssen, Cambridge: Cambridge University Press.
- Smith, M. (1994) *The Moral Problem*, Oxford: Blackwell.
- Smith, M. (1995) "Internalism's Wheel," *Ratio* 8 (new series): 277–302.
- Street, S. (2006) "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127: 109–66.
- Sturgeon, N. (1995) "Moral Explanations," in D. Copp and D. Zimmerman (eds.), *Morality, Normativity, and Society*, New York: Oxford University Press.
- Turiel, E. (2006) "The Development of Morality," in N. Eisenberg (ed.), *Handbook of Child Psychology*, Vol. 3, 6th edition, New York: Wiley, 789–857.
- van Roojen, M. (2006) "Knowing Enough to Disagree," *Oxford Studies in Metaethics* 1: 161–94.