

## Conceptual Analysis in Metaethics

Nicholas Laskowski and Stephen Finlay

### INTRODUCTION

Despite being considered the traditional and prototypical philosophical method, conceptual analysis is an approach to philosophical inquiry with a checkered past and a tarnished reputation. This is particularly true in metaethics, which we understand here as concerned not only with ethics or morality proper but also with the normative more generally. In recent decades, metaethicists have often consigned conceptual analysis to the trash pile of philosophical mistakes, yet it has always had adherents and today enjoys a renaissance. In this selective and opinionated chapter, we explore various dimensions of the debate, framed around two foundational questions: (1) what is it to *use* and *possess* a normative concept? and (2) what is it to *analyze* a normative concept? We attempt to avoid broader questions about the nature and analysis of concepts in general (for orientation, see Margolis & Laurence, 1999, 2015).

*Analysis* we understand, most broadly, as investigation into or explication of something's nature. All we assume at the outset about the nature of *concepts* is (1) that they are psychological resources employed in and enabling us to have *thoughts*, and (2) that they are psychological *types* of some kind rather than tokens—whether of mental representations, contents, abilities, processes, or similar—and therefore that different people can share the same concept. We make no assumptions about what makes a concept *normative*, except that normative concepts are those whose involvement in a thought make that thought normative—as it is the involvement of the concept WRONG that makes the thought that killing is wrong normative, rather than the concept KILLING. The extension of the class of normative concepts is controversial—contested cases include CRUEL, TRUE, and MEANING—but our treatment focuses on less controversially normative concepts such as WRONG, GOOD, OUGHT, and REASON.

The conceptual domain is helpfully juxtaposed against the linguistic and metaphysical domains, which it is commonly viewed as mediating. First, consider the relationship between the conceptual and the linguistic. A common assumption is that many (so-called

*lexical*) concepts can be identified as the conventional meanings of words, a connection we'll call the *word-concept nexus*. Conceptual analysis is therefore often equated with, and pursued by means of, semantic analysis or study of linguistic meaning. This article focuses on such lexical concepts, following the convention of denoting them with corresponding words in small caps; for example, 'WRONG' refers to the conceptual meaning of the English word 'wrong'. For convenience, we will typically speak as if words and concepts stand in a one-to-one relationship, although readers should be wary of this assumption. On one hand, WRONG is potentially also the meaning of the German word 'unrecht' and the Swahili word 'vibaya', for example; on the other hand, some normative words may be ambiguous or context-sensitive in meaning, so it may be necessary to distinguish between WRONG<sub>1</sub>, WRONG<sub>2</sub>, and so on.

Second, consider the relationship between the conceptual and metaphysical domains. The latter potentially includes normative *entities* such as reasons and obligations, normative *properties* such as wrongness and goodness, normative *relations* such as counting-in-favor-of, and normative *states of affairs* such as that lying is sometimes wrong. Philosophical discussion often takes for granted that the basic function of concepts is to "pick out" entities or properties in the world, enabling us to talk and think *about* them; we'll call this connection the *concept-world nexus*. On this basis, conceptual analysis is traditionally thought to provide a path to understanding not only normative concepts and thoughts but also the nature of normative properties and facts. This hoped-for metaphysical payoff is a central motivation of conceptual analysis in metaethics, but, as we will show, normative concepts can also be of interest for their own sake.

### THE CLASSICAL THEORY, PRIMITIVISM, AND THE OPEN QUESTION ARGUMENT

Philosophical tradition presents the Classical Theory of the nature of concepts and conceptual analysis, which traces its (Western) roots back at least to a foundational metaethical text, Plato's *Euthyphro*. Our discussion is organized around this view (following the model of Margolis & Laurence, 1999): after sketching its central tenets, in subsequent sections we introduce more recent, alternative views on normative concepts in terms of their departures from the Classical Theory.

The Classical Theory involves at least the following core tenets:

1. *Cognitivism*: Concepts have an essentially referential or representational function (or "role") of picking out items at the metaphysical level. For example, the concept WRONG enables us to ascribe and think about the property of wrongness, and the concept OBLIGATION enables us to ascribe and think about obligations, as normative entities.
2. *Definitionism*: Concepts fulfill this referential function by providing necessary and sufficient conditions for something's falling within the concept's extension—a set of conditions that may be articulated in the form of a descriptive definition. For example, Socrates' stalking horse in the *Euthyphro* is the proposed definition of PIOUS as *what is approved by the Gods*.
3. *Compositionality*: Concepts can be either complex (structured) or simple (primitive, atomic). Complex concepts are composed out of simpler conceptual parts. For

example, on Euthyphro's view, the concept *PIOUS* is composed at least of the simpler concepts *APPROVED* and *GODS*.

4. *Reductivism*: To analyze a concept is to decompose it, revealing its constituent parts and how they are interrelated—particularly, by providing its descriptive definition.
5. *Analyticity*: Some thoughts and sentences (termed “analytic”) are guaranteed true simply because of the concepts or meanings of the terms involved. On Euthyphro's view, for example, it is analytic that if the gods approve of prosecuting one's father for murder then it is pious to prosecute one's father for murder.
6. *Essentialism*: The internal structure of a concept mirrors the internal metaphysical structure of the normative property (entity, etc.) that it picks out. A corollary: by reductive analysis of a concept, we can also derive a reductive metaphysical analysis of the corresponding property.

Additionally, the Classical Theory accepts the following tenets about the epistemology of concepts:

7. *Competency as Grasping a Definition*: Possession of a concept consists in (some kind of) psychological grasp of its descriptive definition; competence with a word consists in associating it with its definition.
8. *Transparency*: Being mental entities, concepts are transparent to reflection. They therefore can be analyzed, and any analytic truths identified, from the armchair in exercises of purely a priori or nonempirical inquiry (i.e. by “intuition”).

The application of this classical package to normative concepts is exemplified in the early twentieth century in the work of G.E. Moore (1903). However, Moore also sent an early shot across the bow of the ambitions of conceptual analysis in metaethics, independent of any of the objections later raised against the Classical Theory, with a challenge that largely set the stage for subsequent debate.

By the classical tenet of compositionality, all concepts are ultimately made up out of simple or primitive conceptual elements. This raises the possibility that some *normative* concepts are among these simple elements. By the tenet of reductivism, such simple normative concepts would be unanalyzable, since they do not admit of further decomposition or definition. By the tenet of essentialism, this possibility further implies that some normative properties or entities are similarly simple and unanalyzable. This combination of views, which we'll call *primitivism*, is embraced by Moore and many contemporary metaethicists. (It is often labeled “nonnaturalism,” but we'll stay neutral about the correct use of this terminology.) Primitivists disagree over which normative concept or concepts to identify as basic. We'll follow Moore in privileging *GOOD*, although today *REASON* or *COUNTS-IN-FAVOR-OF* are more commonly preferred (e.g. Scanlon, 1998).

Primitivism might seem a relatively minor challenge to the ambitions of conceptual analysis. On Moore's view, for example, every other normative concept is complex and thus analyzable, including *RIGHT* as *PRODUCES MOST GOOD*. Even as the Classical Theory has fallen into general disrepute, few metaethicists have denied that there are at least some illuminating analyses of normative concepts along classical lines, such as the popular reductive analysis of *MURDER* as (at least partly) *WRONGFUL KILLING*. But a guiding aim of metaethical inquiry has been to explicate the puzzling nature of normativity

*per se*, and hence of normative concepts as a *class*. Therefore, metaethicists have typically been interested primarily in the possibility of (conceptual and metaphysical) analysis of the normative at the *global* level rather than of any individual normative concept or property. Most metaethical discussion of conceptual analysis is therefore directed at the prospects for analyzing all or any normative concepts into complexes of entirely nonnormative elements, a Holy Grail widely equated with “analytic reduction” of the normative as such.

Why believe any normative concepts are simple and unanalyzable? The primary motivation is Moore’s famous and still influential *Open Question Argument* (OQA), which relies heavily on the tenets of the Classical Theory. For our purposes, this argument can be presented as follows:

- OQ1. For any description D and concept M, if D is a correct definition or analysis of M, then it must be a *closed question* (roughly, one with a self-evidently positive answer) for competent users of M that whatever is D is also M, and vice versa.
- OQ2. There exists no description D such that it is a closed question for competent users of the concept GOOD that whatever is D is also good, or vice versa.
- OQ3. Therefore, there exists no description D that is a correct definition or analysis of the concept GOOD.
- OQ4. (*Conceptual Conclusion*) Therefore, GOOD is a simple and unanalyzable concept.
- OQ5. (*Metaphysical Conclusion*) Therefore, goodness is a simple and unanalyzable property.

Many metaethicists view Moore’s ultimate, metaphysical conclusion as unacceptable and therefore reject one or another of the argument’s steps or assumptions—which usually involves the rejection of one or more classical tenets. Subsequent sections investigate the major responses in roughly reverse order of how radically they depart from the Classical Theory. We will suggest that more extreme rejections of the Classical Theory may be overreactions and that a subtler departure may be a better course. However, we first conclude this section by observing some problems for the primitivist’s (typically classical) treatment of normative concepts. (We largely ignore primitivist treatments of normative properties, as they fall outside our scope.)

One thorny issue for primitivism derives from a puzzle for the Classical Theory about simple concepts, its basic building blocks. Notice that the tenet of definitionism (Classical Theory’s official story of how concepts refer to properties in the world) as well as the tenet of competency as grasping a definition (the official story of what it is to possess a concept) cannot extend to simple concepts, which as such lack descriptive definitions. Primitivism therefore owes some other account of how simple normative concepts pick out normative properties and of what possession of them consists in. This challenge might not seem especially difficult, since it is a well-established view in general philosophy of mind that some—perhaps even all—concepts are primitive (e.g. Jerry Fodor’s conceptual atomism; for discussion, see Margolis & Laurence, 1999: 59–71). However, metaethical primitivists face a special problem because conceptual atomists generally explain the concept-world nexus by appeal to causal connections between the properties in the world and the concepts in our minds. This answer is not available to most metaethical primitivists, who hold that basic normative properties are not causally efficacious.

A related difficulty concerns how simple normative concepts are acquired if normative properties never causally interact with us or our senses in any way. Here, primitivists might seem obliged to embrace nativism, the Platonic idea that normative concepts are innate. This in turn raises difficult questions about how we could ever be justified in believing that our normative concepts successfully pick out anything that actually exists or that our normative thoughts ever correspond to reality. Metaethicists who have flirted with nativism have indeed tended to be error theorists, denying that normative concepts and thoughts ever correspond to anything actually existing in the world (e.g. Joyce, 2006). In our opinion, primitivists owe these problems of concept reference, possession, and acquisition greater attention than they have received.

Primitivists generally maintain that basic normative truths are a priori and knowable from the armchair by intuition alone—as is widely accepted. This forces an awkward choice. The a priori status and intuitionistic methodology of ethics seem easily explained (by the classical tenet of transparency) if these propositions are *analytic* and therefore knowable through conceptual analysis—but this explanation seems (prima facie) unavailable to primitivism and would apparently falsify OQ2, which denies the existence of any closed questions linking GOOD with a description D in nonnormative terms. The second option is to maintain that the propositions are nonanalytic—or *synthetic*—in which case their a priori status is widely perceived to be mysterious.

Traditionally, primitivists take this second path, embracing (e.g. Enoch, 2011) or denying (e.g. Scanlon, 2014) the mystery. As is seldom noticed, this choice also stands in a potentially embarrassing tension with the OQA. This is because the notion of a “closed question” is perilously close to that of an a priori truth (see further discussion below). Premise OQ2 may therefore prove incompatible with the intuitionist claim that first-order normative propositions are knowable a priori (Soames, 2011). Recently, some primitivists have reached for the first option above, conceding that a priori normative propositions are indeed conceptual truths. They maintain their rejection of analytic reductions of normative concepts by allowing only conditionals and never biconditionals: for example, an action is wrong *if* it is an intentional infliction of suffering on the innocent for fun, but not *only if* (Audi, 2004; Huemer, 2005; Cuneo & Shafer-Landau, 2014). We worry that this strategy may succeed only in moving the bump in the rug. To avoid the result that normative concepts can be reductively analyzed after all, as massive disjunctions of such conditionals, it seems committed to some further (basic) normative element in a concept like WRONG or GOOD, and the possibility of knowing (nonanalytic) truths involving this further element remains unaccounted for.

### NATURALISM, SYNTHETIC DEFINITIONS, AND SEMANTIC EXTERNALISM

These problems seem easily avoided if we reject primitivism in favor of a metaethical naturalism that identifies or reduces normative properties to causally efficacious, “natural” properties. This option is supposed to be ruled out by the OQA’s final inference, from its conceptual to its metaphysical conclusion. However, the inference is fallacious: basic normative *properties* could be complex and analyzable, contra primitivism, even if normative *concepts* aren’t. This amounts to rejecting at least the classical tenet of essentialism, which looks vulnerable: plausibly, not every concept picks out its object by that object’s

essence or basic composition. On most theories, the same objects or properties can be picked out by different concepts; famously, the planet Venus can be picked out by either the concept MORNING STAR or the concept EVENING STAR. So the correct reductive or essential definitions of normative properties like goodness could be *synthetic* rather than analytic, like the definition of water as H<sub>2</sub>O (e.g. Railton, 1986; Copp, 2007). In that case, their discovery wouldn't be possible through a priori conceptual analysis alone and might require empirical investigation. Rejection of essentialism and appeal to synthetic definitions doesn't dispose of all the puzzles about normative concepts, however. For example, it doesn't by itself explain why it would be so difficult to find *any* analytic truths connecting normative and nonnormative terms. Helen and Richard Yetter-Chappell (2013) argue that synthetic naturalists might take inspiration from strategies deployed in the philosophy of mind to explain away "explanatory gaps" left by physicalist definitions of consciousness.

A particularly radical departure from the Classical Theory is found in "Cornell Realism" (e.g. Brink, 1984; Boyd, 1988), which seeks at once to evade the OQA's challenge and put metaphysical analysis of normative properties on a respectably naturalistic footing by freeing it from conceptual analysis altogether. This combines synthetic naturalism with the doctrine of *semantic externalism*, which rejects the classical tenet of competency as grasping a definition (or "description theory"). The 1970s saw the introduction and embrace of the view that making reference to something in speech or thought requires only acquisition of a term linked to the referent by an appropriate causal chain—even if one lacks any individuating description. This can be seen as denying a necessary role for concepts in language and thought altogether (in Hilary Putnam's slogan, "Meaning ain't in the head") and replacing the word-concept nexus and concept-world nexus with a direct *word-world* nexus. The strongest version of such naturalism repudiates not only the highest ambitions but also the very possibility of armchair conceptual analysis, and its dominance over philosophical thought in the 1980s contributed greatly to the disrepute of this methodology.

Over recent years, however, the popularity of this kind of naturalism in metaethics seems to have faded. Observing that normative thought and language could conceivably function this way is one thing; demonstrating they actually function this way is quite another, and the case for synthetic definitions of and causal reference to normative properties seems weak. Not every class of words is as amenable to this treatment as the paradigmatic examples of names ('Aristotle') and natural kind terms ('water', 'tiger')—consider mathematical vocabulary, for example—and it seems a particularly poor fit for normative vocabulary (see Pigden, 2012). A primary problem is the already mentioned intuitionist or a priori epistemology of a robust range of normative truths, which is a primary motivation for primitivists' denials that normative properties could be "naturalistic." Commonly cited examples include both basic first-order normative propositions such as that intentionally inflicting suffering on the innocent for fun is wrong, and propositions about the nature of normative properties such as that they supervene on the nonnormative state of the world. This suggests that basic competence with normative thought involves substantial acquaintance with the essences of normative properties. While many philosophers have followed Immanuel Kant in embracing the possibility of synthetic a priori knowledge, particularly in ethics, this doesn't seem a genuine option for Cornell Realists, since it is implausible that causally efficacious or natural facts and properties are the kinds of things into which we could have a priori insight.

### NONCOGNITIVISM, ATTITUDES, AND INFERENCES

An entirely different and perennially popular way of reacting to the OQA involves rejecting the inference from OQ3 (there is no description that is a correct definition of a basic normative concept) to the conceptual conclusion OQ4 (basic normative concepts are simple and unanalyzable) by way of rejecting the core classical tenet of *cognitivism*, that normative concepts have the essentially referential or representational role in thought of picking out normative properties. This *noncognitivist* approach has the virtue of easily avoiding all the problems observed above for primitivism and synthetic naturalism, as explained below. *Hybrid* approaches postulating a combination of cognitive and noncognitive functions are also popular today, although more often formulated at the linguistic than the conceptual level; these have their own strengths and weaknesses (see Teemu Toppinen's chapter "Hybrid Accounts of Ethical Thought and Talk").

General philosophical discussions of concepts often assume that the essential role of concepts is to pick out or refer to things in the metaphysical domain, such as individual entities (Aristotle), kinds (bachelor, horse), and properties (yellow). However, if taking our cues to the range in the conceptual domain from the variety in the linguistic domain, we should notice that cognitivism seems an unpromising approach to the meaning of many classes of words other than ordinary nouns, adjectives, verbs, and adverbs. Consider logical connectives ('and', 'or', 'if', 'not'), modal and sentential operators ('must', 'might', 'perhaps'), interjections ('hurray!', 'ouch!') and performatives ('hello', 'please'); what *in the world* could they be picking out? Rather than denying that these words express any concepts at all, we might expand our understanding of concepts to include other kinds of contribution to thought. So perhaps the essential role of normative concepts also is something other than—or on a hybrid approach, isn't exhausted by—such (so-called) cognitive functions, as many philosophers have proposed since the 1930s. (However, some early noncognitivist accounts, like A. J. Ayer's emotivist theory that 'wrong' is a device for conveying disapproval and/or influencing others, seem best characterized as claiming there are no normative thoughts or concepts, only normative language.)

Noncognitivism is a radical repudiation of the Classical Theory, because it entails rejection also of most of the other tenets we observed, which presuppose cognitivism. Against the tenet of definitionism, it denies the possibility of capturing the meaning of normative terms by a synonymous or intersubstitutable description providing a set of necessary and sufficient conditions. Simply put, the (cognitive) conceptual role of a description is fundamentally different than the (noncognitive) conceptual role of a normative term. This would explain the truth of OQ3, the lack of correct descriptive definitions of normative concepts, in a way that doesn't license Moore's conceptual conclusion (OQ4) that normative concepts are simple and unanalyzable. This is because noncognitivists also reject the tenets of compositionality and reductivism, endorsing a broader, nonclassical account of what it is to analyze a normative concept. On this account, we analyze a normative concept indirectly (from "sideways on") by describing what someone is doing or like when they possess or use it. Noncognitivism is also well placed to explain how substantive and nonanalytic normative thoughts can have an apparently a priori character, and the intuitionistic epistemology of ethics. Whatever they might be, normative thoughts are not about the obtaining of some (metaphysically objective, mind-independent) state of affairs.

What might conceptual analysis of this kind reveal the noncognitive roles of normative concepts to be? Here, there are many options, but existing views largely fall into two

classes. *Expressivist* theories analyze normative thought in terms of “noncognitive,” motivational attitudes like desires, intentions, and emotions, as opposed to the “cognitive” attitude of (ordinary) belief (e.g. Blackburn, 1998; Gibbard, 2003; see Elisabeth Camp’s chapter “Metaethical Expressivism”). *Inferentialist* theories analyze normative thought in terms of inferences to and from other attitudes that it either psychologically or rationally requires (e.g. Wedgwood, 2007; Chrisman, 2015; see Matthew Chrisman’s chapter “Conceptual Role Accounts of Meaning in Metaethics”). For example, an inferentialist analysis of OUGHT might claim that to use the concept is to rationally commit oneself to certain patterns of intention. (Note that such a theory may reject the possibility of broadly reductive analysis in nonnormative terms, since RATIONAL seems itself a normative concept.)

Noncognitivist views are a popular choice from the smorgasbord of metaethical theories, partly because they emphasize practical aspects of normative thought (e.g. in eliciting action, intention, or emotion) that cognitivist views have struggled to accommodate. However, noncognitivists face some difficult challenges. This includes (1) a *prima facie* problem accounting for recalcitrant characters like *amoralists*, who allegedly have moral thoughts without manifesting the motivational attitudes or inferential dispositions that noncognitivists claim to be constitutive of such thoughts, and (2) the avowals of many people that their normative thoughts are cognitivist, representing metaphysically robust normative states of affairs (“moral facts”). These claims are potentially embarrassing for noncognitivists given the classical tenet of transparency—one element of the Classical Theory they generally don’t reject, since it undergirds their appropriation of (the first step of) the OQA. However, we wish to focus on a different issue.

Notoriously, noncognitivist views are faced with an “embedding” (or “Frege-Geach”) problem about compositionality (see Jack Woods’ chapter “The Frege-Geach Problem”). For illustration, consider the simple and naive form of expressivism that analyzes WRONG directly in terms of *disapproving* of something. This may yield a plausible analysis of thinking that lying is wrong as disapproving of lying, but it seems unable to account for the more complex thought that lying is *not* wrong, which apparently doesn’t involve disapproving of anything. Similarly, the complex thoughts that *if* lying is wrong *then* deceiving is wrong, and that *S believes that* lying is wrong, are presumably uses of the concept WRONG, but they also apparently don’t involve disapproving of anything. While debate often focuses exclusively on these kinds of logically complex thoughts (or sentences), at its most general the embedding problem concerns noncognitivism’s ability to provide an analysis of a concept or word’s meaning that accounts for the full range of different ways the concept or word can be used or deployed. This includes, for example, explaining what it is to *wonder* whether lying is wrong or to *suppose* or *hope* that lying is wrong, none of which apparently involve disapproving of anything. More sophisticated (and recent) noncognitivist theories are designed with the aim of solving the embedding problem, at least with respect to logically complex thoughts—but despite decades of debate there is little consensus over whether it can be solved or even what a solution must accomplish, and noncognitivists have barely attempted to address it for wondering and other non-belief-like thoughts. Here we focus on two aspects of the problem that have largely been overlooked but that emerge from thinking particularly about normative concepts rather than sentences or thoughts.

First, despite apparently disowning simple views of what it is to *deploy* a normative concept, noncognitivists still sometimes endorse a correspondingly simple view of the

normative concept *possession*: that to possess a concept like *WRONG* is just to be capable of having the associated attitude, like disapproval or blame (e.g. Schroeder, 2008). Given the range of permissible uses of *WRONG*, this view makes concept possession come apart from concept deployment in an awkward way. It isn't obviously incoherent to imagine a person thinking that lying is not wrong, despite her being psychologically incapable of disapproval or blame. According to the suggested view, however, she would apparently be using a concept that she doesn't possess! Noncognitivists must either challenge the possibility as characterized or provide a different account of concept possession.

Second, while noncognitivists have focused on solving the embedding problem at the *sentential* level, or the level of embeddable normative thoughts such as that lying is wrong, even if these efforts succeed it can be questioned whether noncognitivism offers a viable *alternative* to the Classical Theory at the level of individual normative concepts like *WRONG*. To illustrate this, consider an expressivist theory designed expressly to address the embedding problem, *biforcated attitude semantics* or BAS (Schroeder, 2008). This attempts to solve the problem by splitting the attitudinal component of normative thought into two parts, identifying the attitude of belief as a generic pro-attitude of "being for" taken towards other mental states. BAS analyzes the belief that lying is wrong as the state of mind of *being for blaming for lying*. By splitting the attitudinal component, BAS creates space for logical operators like *not* and *if* in the thought's content. For example, the belief that lying is not wrong is analyzed as *being for not blaming for lying*. Unlike simple expressivist theories, then, BAS can apparently identify a common contribution 'wrong' makes to the states of mind expressed by the sentences 'Lying is wrong' and 'Lying is not wrong': in each case, it contributes the concept *BLAMING FOR*.

Notice, however, that the attitude-type of blaming for is not *itself* involved in thinking that lying is not wrong (nor, indeed, in thinking that lying is wrong!). Rather, the thought is analyzed as an attitude *towards/about* the state of mind, *not blaming for lying*. BAS thus assigns the concept *WRONG* a traditionally "cognitive" role (of picking out a particular type of psychological attitude), the same *kind* of role as (on a classically cognitivist theory) the concept *LYING*. The noncognitive character of the thought expressed by asserting 'Lying is wrong' is contributed by the noncognitivist treatment of belief itself. (BAS recovers a distinction between cognitive and noncognitive thoughts by offering a radically nonclassical analysis of the former.) Schroeder (2008) claims, controversially, that the embedding problem can only be solved by a noncognitivist theory with this split-attitude structure. If this is correct, noncognitivism may fail to provide a viable distinctive approach to normative concepts like *WRONG*, which must make a classically cognitive contribution to normative thoughts to support the full range of ways we use them. To be clear, this is not an objection to noncognitivism at the level of whole normative thoughts, but we suggest it is an unappreciated obstacle to conceiving of noncognitivism as a distinct option at the lexical level of individual normative concepts, as it is being considered here.

## DEFINITIONS, PROTOTYPES, AND NETWORKS

This section presents two different approaches to normative concepts and their analysis—Prototype Theory and Network Theory—that depart from the Classical Theory in

a less radical way, accepting the central tenet of cognitivism but rejecting the tenet of definitionism, at least as classically understood.

Prototype Theory is a broad family of views tracing their lineage back at least to Ludwig Wittgenstein's discussion of family resemblance. Concepts are understood in terms of a prototype or exemplar that a subject has in mind, which applies to things depending on the degree to which they approximate the prototype as *sufficiently like THAT*. For a (classic) example, one's mental prototype for GAME might be soccer, in which case a central application could be to rugby (conventional rules, a team sport, involves a ball, competitive, involves scoring points, played on a field) and less centrally to chess (conventional rules, competitive, but doesn't involve teams, balls, a field, or scoring points). This rejects definitionism's reliance on necessary and sufficient conditions as the criteria of concept-application. Perhaps no single feature is *necessary* for something to fall under a concept like GAME: games needn't have conventional rules, for instance, or be competitive. Other classical tenets that presuppose definitionism are collateral damage. Contra compositionality, concepts are not constructed out of simpler parts, and so, contra reductivism, they are not analyzed by decomposition. Conceptual analysis instead takes the form of identifying the prototype or listing and weighting the features relevant to falling under the concept. There might be no analytic truths involving the concept (other than, perhaps, applications to the prototype itself), since the concept can in principle apply in the absence of any paradigmatic feature provided enough others are present. This supplies a potential explanation for OQ2, the lack of closed questions connecting basic normative properties with nonnormative features (Goldman, 1993). Competency isn't grasping a descriptive definition but the ability to represent and compare the prototype.

Although Prototype Theory is an important and influential competitor to the Classical Theory in the general theory of concepts, in metaethics it lies outside the mainstream. (Proponents include Stich, 1993; Wong, 2006; Park, 2013.) To hazard an explanation, while it seems easy to imagine exemplars for a central normative concept like WRONG, philosophers remain unable to agree on the relevant features of even paradigmatically wrong actions, like inflicting suffering for fun. Prototype Theory seems more plausible and popular for so-called *thick* normative concepts like CRUEL—but the thicker the concept, the more controversial its classification as “normative” becomes. (For discussion of the strengths and weaknesses of Prototype Theory in general, see Margolis & Laurence, 1999: 27–43.)

By contrast, Network Theory, as championed particularly by Frank Jackson (1998; see also Jackson & Pettit, 1995), has made significant noise in metaethics. This approach is often characterized as a return to traditional conceptual analysis, and indeed Jackson is occasionally represented, erroneously, as virtually a lone voice defending conceptual analysis in the metaethical wilderness. However, Network Theory seems well classified as an application of what is known in the general theory of concepts as the Theory Theory (for an overview, see Margolis & Laurence, 1999: 43–51) and thereby is better understood as a modest departure from the Classical Theory.

Network Theory treats normative concepts or terms on the model of the theoretical terms of a scientific theory, with meanings or definitions determined not individually but rather holistically by their role or position in the overall theory, or *network*. It can be roughly but usefully understood as the result of applying Classical Theory directly to the total network rather than to individual normative concepts. In Jackson's hands, the

relevant network is that of *fully matured folk moral theory* and consists in three kinds of connections, which have the status of constitutive but revisable “platitudes”: (1) internal connections between different normative concepts (e.g. that *WRONG* applies to an action just in case *CONCLUSIVE REASON AGAINST* applies to it), (2) “input” platitudes connecting observable states of affairs with normative concepts (e.g. if an action involves inflicting suffering for fun, then it falls under *WRONG*), and (3) “output” platitudes connecting normative concepts with various responsive behaviors (e.g. aversion in response to applications of *WRONG*) (1998).

The departures from the Classical Theory are significant. Network Theory may not even recognize the existence of concepts at the lexical or individual level, the conceptual unit instead being the network. Although Jackson seems to invoke individual concepts, like *WRONG*, he explains that this is just a way of speaking about words (1998: 33). The conceptual and referential role of words like ‘wrong’ is determined by their relations to other parts of the theory, as revealed by the various platitudes. Contra (classical) definitionism, it is the theory that provides necessary and sufficient conditions—for the application of the entire network. Against the tenets of compositionality and reductivism, conceptual analysis proceeds not by decomposing complex individual concepts into simpler parts but by assembling the relevant platitudes (by pumping intuitions from the armchair) into a map of the network, taking the form of an extremely complex description. While Jackson understands conceptual analysis as an exercise in global reduction, translating a description of reality in one (e.g. normative) vocabulary into a discontinuous (e.g. scientific) vocabulary, the actual reductive work is assigned to a separate, empirical inquiry into which overall constellation of entities, properties, and relations at the metaphysical level contingently manifests (close enough to) the same relations. So contra the tenet of essentialism, the essential or reductive definitions of normative properties are synthetic rather than analytic. This provides one part of Jackson’s strategy for accommodating OQ2, the lack of analytic connections between normative and nonnormative terms; additionally, the platitudes themselves are not fully analytic or guaranteed true, since each individually is potentially defeasible under the pressure towards the mature or ideal folk theory.

A general difficulty confronting Network Theory is a *Permutation Problem*: there is no clear guarantee that the complex description of the network will correspond uniquely to one complex of properties, in which case the theory underdetermines references for normative terms (Smith, 1994). We suggest additionally that these departures from the Classical Theory intuitively put the cart before the horse. Whereas Network Theory suggests that we are able to recognize wrongness only derivatively, through our knowledge of platitudes such as that inflicting suffering for fun is wrong, it seems more natural to suppose instead that we know the truth of such “platitudes” because we are able immediately to recognize wrongness (when we see it) much of the time.

### CHALLENGING THE PREMISES OF OQA: NONTRANSPARENCY AND EMPIRICAL LINGUISTICS

We have examined a number of ways in which ambitious conceptual analysis has been defended against the OQA by rejecting one or another core tenet of the Classical Theory

as applied to normative concepts. In this section, we consider whether these responses might be overreactions, because we should instead reject the initial premises of Moore's argument. Consider first OQ2: is it really impossible to provide reductive definitions of basic normative words or concepts that yield "closed questions"? A very few philosophers have been bold enough to reject the premise as demonstrably false: for example, Paul Ziff (1960) proposes a reductive analysis of 'good' as meaning *answers to the interests in question*. No such proposal has won widespread acknowledgment, but the premise can be resisted without being refuted. How such a bold premise can be justifiably *accepted* is unclear when it is impossible to consider every possible definition of 'good'. Perhaps the correct analysis has simply yet to be considered; then what we need is redoubled effort at classical conceptual analysis rather than its abandonment.

A modest approach seeks to support OQ2 by induction: we've tried enough potential conceptual analyses of 'good' without yielding closed questions that we can safely infer that anything we could try will similarly fail. This seems an exceptionally weak induction, however, moving from a small sample to a conclusion about all possible definitions. Another approach proposes an abductive justification: unanalyzability is the *best explanation* for why efforts at conceptual analysis of 'good' have been unsuccessful so far (e.g. Ridge, 2014). But this has yet to be demonstrated; perhaps GOOD is just very complex, for example.

The most ambitious defense of OQ2 involves a particularly strong interpretation of the classical tenet of transparency. Perhaps our normative concepts are so transparent that we can immediately recognize not only that a proposed definition is correct when it is but also what the correct definition or composition of a concept is. Therefore, we can directly perceive, in a "flash of light" (Wittgenstein, 1997) or a "just-too-different intuition" (Enoch, 2011), that our normative concepts are unanalyzable in nonnormative terms. However, even among primitivists, there is widespread disagreement over which normative concepts are basic and over the correct (nonreductive or normative) definitions of others. This stance therefore requires a difficult balancing act, finding sufficient transparency to license the just-too-different intuitions but not so much as to predict that definitional connections between normative concepts would be obvious.

These interpretative issues about the tenet of transparency also point towards an important challenge to premise OQ1, that correct analyses always produce closed questions (having self-evident positive answers). Ambitiously reductive conceptual analysis can be defended against the OQA by either abandoning or weakening the classical tenets of transparency and competency as grasping a definition. This is a minimal departure from the Classical Theory, because it is compatible with retaining all the core (non-epistemological) tenets concerning the nature of normative concepts and conceptual analysis: cognitivism, definitionism, compositionality, reductivism, analyticity, and essentialism.

Just how transparent must our own concepts be to us? At one extreme is a view we'll call *strong transparency* (or strong apriority): that nobody could fail to know the analysis of their own concept upon reflection. This may follow if "grasping" a concept is simply mentally tokening its descriptive definition. However, the psychological link between concept and definition may be interpreted in subtler ways. At the other extreme is the view that our own concepts are entirely opaque to us. Perhaps our minds are so compartmentalized that the resources employed in first-order thought about the world are impenetrable to conscious introspection. In this case, a normative proposition could

be analytically true without being a priori in even the most attenuated sense. Between these extremes lie many possibilities we can collectively call *weak transparency* (or weak apriority). Perhaps possessing a concept merely entails a disposition to recognize the correct definition when presented with it? Or perhaps, more weakly still, possessing a concept merely entails that it is *possible* (although perhaps extremely difficult, requiring advanced philosophical skills and extensive focus) to identify correct definitions and analytic truths.

For the widely accepted first step of the OQA to be both sound and compelling, a fairly strong form of transparency is required. But strong forms of transparency lead to the so-called Paradox of Analysis, as follows: if a conceptual analysis is correct then it must be uninformative because already known, and if it is informative because not already known then it must be incorrect. Since some conceptual analyses are plausibly both informative and correct, there seem to be some *unobvious analyticities* (see King, 1998). David Lewis (1989) exploits this point to defend a variant of the very analysis Moore originally attacked, that GOOD is the concept WHAT WE DESIRE TO DESIRE.

If we reject the tenet of transparency altogether, by what method can conceptual analysis be conducted? Or if we accept only a weak version of transparency, how can we adjudicate between competing analyses or claims to conceptual expertise? One answer is provided by the approach of contemporary linguistic semantics, in the tradition of Ordinary Language philosophy. Although not typically characterized as analysis of “concepts,” this employs methods that, although neutral with regard to the theories we’ve surveyed, are at least compatible with a minimal departure from the Classical Theory. Assuming the word-concept nexus, *lexical* concepts are ripe for investigation by abductive and scientific methods. Different hypotheses about the conceptual meanings of normative words yield differing, testable predictions about competent speakers’ dispositions to use them and assent to their use. The primary data here include speakers’ intuitions about the acceptability of individual sentences incorporating these words, but this can be supplemented with usage, etymological, and cross-linguistic data. Speakers’ intuitions can be gathered by fieldwork (such as the surveys favored by experimental philosophy), but the most common methodology is for the researcher to mine her own linguistic intuitions as a competent speaker herself.

The appeal here to intuitions might be thought at odds with our characterization of this method as empirical rather than a priori. But although practiced from the armchair, it is a process of abductive reasoning from observed phenomena (our introspected “gut reactions” to individual uses of a word) to underlying and hidden causes (our implicit concepts). This approach to normative language has made great progress in linguistics since the 1970s—especially in the work of Angelika Kratzer (1981) on the meaning of modal verbs like ‘ought’ and ‘must’, which was largely overlooked by metaethicists until very recently. However, similarly abductive linguistic methods have also been effectively practiced in metaethics for decades, as in the analysis of GOOD as a predicate modifier on the basis of identifying ‘good’ as an *attributive adjective* (e.g. Geach, 1956; Foot, 2001; Thomson, 2008).

One distinctive break from more traditional conceptions of conceptual analysis is that because this method is self-consciously linguistic, it seeks to identify the full range of different sentential and conversational contexts in which a particular word like ‘ought’ or ‘good’ is used rather than looking narrowly at the distinctly moral or (slightly less

narrowly) normative uses of particular metaethical interest. For example, a sentence like ‘Ben ought to be at work’ allows two very different kinds of readings, one normative and the other epistemic or predictive. Ignoring the latter might seem appropriate given the metaethicist’s interest in *normative* concepts, but if the same lexical concept OUGHT occurs in each reading then this is to ignore potentially important clues to its meaning. One might think that since the two readings involve different thoughts, the word ‘ought’ must be ambiguous between at least two concepts, OUGHT<sub>NORMATIVE</sub> and OUGHT<sub>EPISTEMIC</sub>. But not every element of our thoughts is explicitly represented in the grammar of the sentences that express them. Linguists and philosophers have successfully developed unifying *contextualist* theories of meaning for words like ‘ought’ by identifying other, usually implicit elements that vary between different uses of the same sentence (see Alex Silk’s chapter “Metaethical Contextualism”). The differences between the concepts OUGHT<sub>NORMATIVE</sub> and OUGHT<sub>EPISTEMIC</sub> are then located not in the (unified) *lexical* concept OUGHT but in contextually supplied parameters that aren’t explicit in the sentence. This contextualist approach also offers a further response to the OQA (e.g. Prior, 1964; Foot, 2001; Thomson, 2008). Perhaps the reason why any proposed reductive definition of a normative term fails to generate closed questions is that for every definition that correctly analyzes *one* way of supplementing the lexical concept, there are indefinitely many other ways of supplementing that concept to reach a complete thought or predicate; for example, even if ‘good’ isn’t lexically ambiguous, ‘Is it good?’ may be a radically ambiguous question.

This neo-classical approach to conceptual analysis faces important challenges, but its proponents are not without replies. One challenge is that because words can be lexically ambiguous—like ‘bank’, which expresses at least the concepts BANK<sub>RIVER</sub> and BANK<sub>FINANCIAL</sub>—a linguistic methodology faces the risk of wild goose chases after nonexistent unifying concepts. However, we believe that the track record of this approach warrants considerable optimism (for a sustained case, see Finlay, 2014). Another challenge is the possibility that different people use the same words with different concepts—a threat to take seriously given divergences in people’s moral intuitions. However, moral disputes seem pre-theoretically like substantive disagreements rather than cases of talking past one another, suggesting common concepts with disputed application. (For discussion of the problem of normative concept identity and a novel solution, see Schroeter & Schroeter, 2014.) A related complaint is that it is objectionably parochial to focus on English words like ‘good’ and ‘ought’ or to assume that speakers of other languages employ the same concepts in their normative thought. However, plausibly at least “thin” normative concepts like those expressed by ‘good’, ‘wrong’, and ‘ought’ are practically universal in human thought, since interpreters seldom hesitate to offer these words as translations of central words in other languages. Additionally, if a society were to conduct its normative thought with different concepts it is unclear what bearing this would have on our metaethical questions, which presumably are couched in our own concepts. Rather, such a discovery would just bring to light new questions we hadn’t previously entertained.

Finally, it is complained that the normative concepts most relevant for philosophy aren’t necessarily those that happen to be lexicalized in natural language. A rival model of philosophical inquiry enjoins abandoning conceptual analysis in favor of “reforming definition” (Brandt, 1979), addressed to the question of what concepts we ought to use our words to express. But determining which concepts we ought to use plausibly requires an understanding of the concepts we’re already using (Plunkett, 2016), which

is what neo-classical conceptual analysis seeks to provide. Since metaethicists' interests are plausibly continuous with those of the "folk," and metaethical questions plausibly arise out of reflection on ordinary normative thought and speech, it is also far from clear that the concepts of metaethical significance aren't just those expressed by our everyday normative vocabulary.

In conclusion, although conceptual analysis remains a controversial approach to metaethics, we suggest there are many more conceptions or varieties of it than is commonly recognized, each of which supports an active research program with devoted champions.

### ACKNOWLEDGMENTS

We thank James Fritz, Tristram McPherson, David Plunkett, and Mark Schroeder for helpful comments.

### REFERENCES

- Audi, R. (2004) *The Good in the Right*, Princeton, NJ: Princeton University Press.
- Blackburn, S. (1998) *Ruling Passions*, Oxford: Oxford University Press.
- Boyd, R. (1988) "How to Be a Moral Realist," in G. Sayre-McCord (ed.) *Essays on Moral Realism*, Ithaca, NY: Cornell University Press: 181–228.
- Brandt, R. (1979) *A Theory of the Good and the Right*, Oxford: Oxford University Press.
- Brink, D. (1984) "Moral Realism and the Skeptical Arguments from Disagreement and Queerness," *Australasian Journal of Philosophy* 62: 111–25.
- Chrisman, M. (2015) *The Meaning of "Ought"*, Oxford: Oxford University Press.
- Copp, D. (2007) *Morality in a Natural World*, Cambridge: Cambridge University Press.
- Cuneo, T. and Shafer-Landau, R. (2014) "The Moral Fixed Points: New Direction for Moral Nonnaturalism," *Philosophical Studies* 17(3): 399–443.
- Enoch, D. (2011) *Taking Morality Seriously*, Oxford: Oxford University Press.
- Finlay, S. (2014) *Confusion of Tongues*, Oxford: Oxford University Press.
- Foot, P. (2001) *Natural Goodness*, Oxford: Oxford University Press.
- Geach, P. (1956) "Good and Evil," *Analysis* 17(2): 33–42.
- Gibbard, A. (2003) *Thinking How to Live*, Cambridge, MA: Harvard University Press.
- Goldman, A. (1993) "Ethics and Cognitive Science," *Ethics* 103: 337–60.
- Huemer, M. (2005) *Ethical Intuitionism*, Houndmills: Palgrave MacMillan.
- Jackson, F. (1998) *From Metaphysics to Ethics*, Oxford: Oxford University Press.
- Jackson, F. & Pettit, P. (1995) "Moral Functionalism and Moral Motivation," *Philosophical Quarterly* 45(178): 20–40.
- Joyce, R. (2006) *The Evolution of Morality*, Cambridge, MA: MIT Press.
- King, J. (1998) "What Is a Philosophical Analysis?" *Philosophical Studies* 90(2): 155–79.
- Kratzer, A. (1981) "The Notional Category of Modality," in H. J. Eikmeyer & H. Rieser (eds.) *Words, Worlds, and Contexts*, Berlin: De Gruyter: 38–74.
- Lewis, D. (1989) "Dispositional Theories of Value," *Proceedings of the Aristotelian Society* 63: 113–37.
- Margolis, E. & Laurence, S. (eds.) (1999) *Concepts: Core Readings*, Cambridge, MA: MIT Press.
- Margolis, E. & Laurence, S. (eds.) (2015) *The Conceptual Mind*, Cambridge, MA: MIT Press.
- Moore, G. E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press.
- Park, J. J. (2013) "Prototypes, Exemplars, and Theoretical & Applied Ethics," *Neuroethics*, DOI:10.1007/s12152-011-9106-8.
- Pigden, C. (2012) "Identifying Goodness," *Australasian Journal of Philosophy* 90(1): 93–109.
- Plunkett, D. (2016) "Conceptual History, Conceptual Ethics, and the Aims of Inquiry," *Ergo* 3(2): 27–64.
- Prior, A. N. (1964) *Logic and the Basis of Ethics*, Oxford: Clarendon Press.

- Railton, P. (1986) "Moral Realism," *Philosophical Review* 95(2): 163–207.
- Ridge, M. (2014) "Moral Nonnaturalism," *Stanford Encyclopedia of Philosophy*. Available at: <https://plato.stanford.edu/entries/moral-non-naturalism/> [Accessed 16 February 2017].
- Scanlon, T. M. (1998) *What We Owe to Each Other*, Cambridge, MA: Belknap Press.
- Scanlon, T. M. (2014) *Being Realistic About Reasons*, Oxford: Oxford University Press.
- Schroeder, M. (2008) *Being For*, Oxford: Oxford University Press.
- Schroeter, L. & Schroeter, F. (2014) "Normative Concepts: A Connectedness Model," *Philosophers' Imprint* 14(25): 1–26.
- Smith, M. (1994) *The Moral Problem*, London: Blackwell.
- Soames, S. (2011) *Philosophical Analysis in the Twentieth Century, Volume 1*, Princeton, NJ: Princeton University Press.
- Stich, S. (1993) "Moral Philosophy and Mental Representation," in M. Hechter, L. Nadel, & R. Michod (eds.) *The Origin of Value*, Berlin: De Gruyter: 215–28.
- Thomson, J. J. (2008) *Normativity*, Chicago, IL: Open Court.
- Wedgwood, R. (2007) *The Nature of Normativity*, Oxford: Oxford University Press.
- Wittgenstein, L. (1997) "Lecture on Ethics," in S. Darwall, A. Gibbard, & P. Railton (eds.) *Moral Discourse and Practice*, Oxford: Oxford University Press.
- Wong, D. (2006) *Natural Moralities*, Oxford: Oxford University Press.
- Yetter-Chappell, H. & Yetter-Chappell, R. (2013) "Mind-Body Meets Metaethics: A Moral Concept Strategy," *Philosophical Studies* 165(3): 865–78.
- Ziff, P. (1960) *Semantic Analysis*, Ithaca, NY: Cornell University Press.