Metasemantics and Metaethics

Laura Schroeter and François Schroeter

Metaethicists disagree about the semantic content of normative and evaluative terms. According to traditional metaethical realists, the semantic role of normative terms is similar to that of natural kind terms: adjectives like 'morally right' pick out a property as their reference and all competent users of the expression co-refer. Error theorists and fictionalists agree that the semantic role of normative terms is to pick out a property, but doubt that any such property is instantiated. Contextualists hold that the reference of normative terms varies according to the context of utterance; relativists hold that the truth of normative claims should be evaluated with respect to the standards of an interpreter; and expressivists maintain that the semantic role of normative terms is to express the motivational states of the speaker. Similar positions have been developed for evaluative terms like 'morally good' or 'brave'. For ease of exposition, we'll focus primarily on normative terms in this chapter, but the issues discussed generalize to evaluative terms.

To adjudicate such disagreements, we need to ask how the semantic content of linguistic expressions in general gets determined. What makes it the case that certain words (and the thoughts they express) have the semantic contents they do? A metasemantic theory seeks to answer this question. Ultimately, metasemantics should explain the relationship between a *psychological state* (competence with the meaning of a normative term) and a *semantic content* (the contribution made by that word to the correct interpretation of the utterance in which it occurs). For instance, if proper names refer to individuals, then what makes it the case that competence with the meaning of a name like 'Aristotle' guarantees that you refer to a specific individual on every occasion of use? To answer such questions, a metasemantic theory must explain which facts about the subject's use of a name are relevant to fixing the reference (e.g. facts about her internal cognitive organization, her inferential or recognitional dispositions, her history of worldly interactions, her social interactions with others, etc.) and how those facts suffice to single out a specific individual as the reference regardless of variation in background beliefs and empirical facts about the context of use. More generally, a metasemantics must explain how the states that constitute semantic competence ground the assignment of specific semantic contents.

There are a wide range of views in the literature about which specific factors may be relevant to determining semantic contents of expressions – including, for example, facts about the subject's cognitive and motivational dispositions, sociological facts about her linguistic community, facts about her physical environment, anthropological facts about human nature, historical facts about natural selection, or irreducible normative facts about morality or rationality (assuming there are any). But an account that reduces meaning facts to, say, sociological facts still needs to go on to explain how those facts are linked to *this particular utterance* if it's to ground an interpretation of that utterance. The same goes for dispositional, anthropological, historical, or other factors that might figure in a full account of how contents are determined – the relevant factors must all be linked to token utterances to explain how the semantic content of the utterance is fixed. So metasemantic theories can all be thought of as providing competing accounts of the relation in virtue of which particular psychological states – like a competent use of a term or a thought deploying a particular concept – have their semantic contents.

In the case of normative terms, we can formulate the metasemantic challenge as follows:

> How does the state that constitutes competence with the meaning of a normative term guarantee that any competent use of that term has semantic content X?

All metaethical positions assume that normative terms have specific semantic contents, so they all assume that a correct metasemantics would vindicate their preferred account of semantic content.

Answering this metasemantic question has been seen as a particularly difficult challenge for traditional normative realists. Because they hold that all competent speakers pick out the very same property with their use of normative terms, traditional realists must explain how such co-reference is possible given persistent disagreement about the applicability conditions of normative terms. On the face of it, there seems to be much wider scope for disagreement about the applicability conditions for 'right' than for such terms as 'chair' or 'water'. Radical and persistent disagreement about which things are chairs would normally lead us to conclude that we were not using words to pick out the same property. So the worry is that we should draw a similar conclusion in the normative case. If there is no plausible way to explain co-reference in the face of normative disagreement (Björnsson 2012), this would be a good reason to favor other metaethical positions that can explain how sameness of semantic content is guaranteed by speakers' competence. More generally, the challenge for all metaethicists is to show how semantic contents can be settled despite non-convergence among competent speakers' understanding and use of normative terms.

In this chapter, we'll survey different metasemantic approaches to explaining how the content of normative terms is determined. We divide the accounts into two broad groups: internalist and externalist approaches diverge over what's required for semantic competence and how facts about competence determine semantic contents. Since the non-convergence problem for context-invariant realism has received the most critical attention in the literature, we will use that position as our primary example for illustrating different metasemantic approaches. But along the way, we'll highlight how the issues we discuss generalize to other metaethical positions. In the final section, we focus on recent developments in metaethics that bring a more nuanced understanding of the role of compositional semantic theorizing in linguistics into the metaethical debate. Metaethicists have traditionally assumed a straightforward relationship between the content of normative and evaluative terms and the content of the thoughts those terms are used to express. But a theory of the compositional semantics for language is not equivalent to a theory of thought content, and recently some theorists have argued that this insight can help to counter one of the main challenges to expressivism, the Frege-Geach problem.

1. The metasemantic task

A metasemantic theory should explain both (i) what constitutes competence with the meaning of a particular expression and (ii) how meeting those competence conditions ensures that the use of that expression has a particular semantic content (see (Peacocke 1992) for this two-part division of labor).

We'll assume that there are genuine semantic facts of both these types. In doing so, we set aside both skepticism about the determinacy of semantic content (e.g. Kripke 1982) and skepticism about individuating fine-grained meanings (e.g. Quine 1951; Stalnaker 2008). This semantic realism is the default assumption within mainstream metaethics.

It's worth saying a few words about how *meaning, semantic competence,* and *semantic content* are related. A semantic content is the contribution made by a word to the correct interpretation of an utterance in which it occurs. For instance, a name like 'Hesperus' may contribute a reference, Venus, to the truth-conditions of utterances in which it occurs, or it may contribute a contextually restricted variable, or a descriptive condition. We'd like to stress that semantic contents in this sense need not be referential – they can be context-dependent functions or markers of expressive significance.

On our usage, meaning is a finer-grained notion than semantic content. For instance, 'Hesperus' and 'Phosphorus' may share the same semantic content, yet differ in meaning. This fine-grained notion of meaning is grounded in (broadly psychological) *ways* of grasping an expression's semantic content. We call these ways of grasping a given semantic content 'competence conditions'. Thus, token uses of 'Hesperus' and 'Phosphorus' differ in meaning just in case they are governed by distinct competence conditions. Conversely, when token uses of an expression are governed by the same competence conditions (as in 'Hesperus = Hesperus'), those uses are guaranteed to have the same semantic content – since particular ways of grasping a content must determine which content is grasped.

Now let's consider our dual metasemantic task. Suppose that 'x is nuba' is a normative predicate in a foreign language. The first metasemantic question is:

> <u>Semantic competence:</u> In virtue of what does someone count as competent with the meaning of this term?

To fully answer this question, we must specify *competence conditions*: particular psychological facts about the use of an expression (perhaps together with non-psychological background facts) that are necessary and sufficient for competence with the standard meaning of 'nuba'. Notice that these conditions must be *specific enough* to distinguish competence with the meaning of 'nuba' from competence with any other possible meaning. In effect, a theory of semantic competence individuates distinct ways of grasping particular semantic contents. An account of competence with specific expressions like 'nuba', moreover, should fit within a general theory of semantic competence with arbitrary expressions in a language.

'Nuba', we are supposing, has certain semantic properties: it is either representational or expressive, context-sensitive or context-neutral, etc. Within these general categories, moreover, we can distinguish the specific content expressed by 'nuba' from that of other terms with similar representational, expressive or context-sensitive contents. So the second metasemantic question is:

> <u>Semantic determination</u>: In virtue of what do competent uses of 'nuba' have these semantic properties?

To fully answer this question, we need a *determination theory,* which explains how specific psychological or non-psychological facts about the use of an expression are relevant to determining its semantic content. Once again, the determination theory for a particular expression like ‘nuba’ must fit within a general determination theory for language and thought content. The task of the determination theory is complicated by the fact that there are indefinitely many different natural, functional and gerrymandered properties and indefinitely many of them will have significantly overlapping instantiation conditions. Similarly, there are potentially infinitely many distinct psychological state types that could be expressed, and infinitely many functions from contexts to extensions. Explaining why just one of these semantic contents is the one expressed by ‘x is nuba’ is a highly non-trivial task.

We take it as common ground in metaethics that competence with the same meaning guarantees sameness of semantic content: anyone who satisfies the competence conditions for ‘x is nuba’ expresses the same content, regardless of how their background beliefs, motivations, or social and physical circumstances may vary. Even relativists and contextualists hold that a particular normative term like ‘nuba’ will always contribute the very same function from contexts of use and standards of assessment to extensions.

This requirement that semantic competence secure sameness of semantic content places an important constraint on the relationship between a theory of competence and a determination theory:

> Linking constraint: The determination theory must appeal *only* to facts about a subject’s understanding or circumstances that figure in the competence conditions for particular meanings.

Suppose a specific belief or disposition is not part of the necessary and sufficient conditions for competence with the meaning of ‘nuba’. Obviously, the determination theory for ‘nuba’ cannot appeal to this belief or disposition to explain how all competent subjects – including those who lack the belief or disposition in question – share the same semantic content. This means that competence conditions must be rich enough to single out a determinate content. At the same time, however, competence conditions must not build in too many substantive constraints on understanding if we are to allow for shared meanings over time and between subjects.

Metaethicists have traditionally assumed there is a straightforward relationship between the contents of normative terms and the contents of the thoughts those terms express. One natural picture is that the meaning of the adjective ‘morally right’ is partly constituted by its connection to a corresponding concept, [morally right], which is what determines the distinctive semantic content expressed by that adjective (e.g. the property of *moral rightness* or the attitude of *moral approval).* There are different ways of conceiving the relation between the lexical entry for an expression and the corresponding concept it expresses. On one approach, lexical entries involve formal semantic “packaging” that determines how an expression will contribute to compositional semantics in a language and “pointers” to specific concepts that lie outside of the language faculty (Glanzberg 2014; Pietroski 2010). For instance, the lexical entry for ‘is tall’ might point to the concept of *height* packaged with a place-holder for a *scale* and a *cut-off point*, which can be supplied by context. This approach ties lexical semantics to the thought contents literally expressed, while allowing for the possibility of linguistic context dependence. See (Väyrynen 2013) for an application to thick concepts. For a purely formal conception of linguistic meaning that leaves out any tie to the thought contents expressed by uses of sentences, see

(Yalcin 2014). As we'll see in §4, this distinction between formal compositional semantics and the substantive thought contents expressed is important to contemporary debates about expressivism.

2. Internalist theories

The traditional approach to the dual metasemantic task is an internalist one:

> Internalism: Semantic competence and semantic contents are determined exclusively by physical factors inside the individual or psychological resources available to the individual.

The intuitive idea is that subjects implicitly know (or have a priori access to) the semantic contents of their own words and thoughts. On standard internalist accounts, to count as competent with a given meaning, one must rely on *rules* or *criteria* for the correct use of an expression. The semantic content of the expression is then settled by those criteria, independently of facts about the individual's external circumstances such as facts about: (a) her own past states, (b) her linguistic community, (c) the metaphysical nature of features of her environment, or (d) causal relations linking the speaker to (a)-(c).

2.1 from simple descriptivism to neo-descriptivism

Simple descriptivism is a paradigm internalist metasemantics. According to simple descriptivism, competence with the meaning of a name or predicate consists in an (implicit) understanding of its applicability conditions in any possible circumstance. To count as competent with the name 'Aristotle', for instance, you must be disposed to rely on a specific criterion that serves as your ultimate standard for identifying what the name applies to. The name's semantic content is simply read off of this internal criterion of application. So the content of 'Aristotle' will be a descriptive condition specifiable by a definite description like 'the last great philosopher of antiquity'. Since competence with the same meaning requires reliance on the same criterion and the criterion fully determines semantic content, it's clear why anyone who meets the competence conditions is guaranteed to pick out the same content. And the linking constraint will be satisfied: semantic assignments depend exclusively on the internal states that constitute semantic competence.

Simple descriptivism, however, came under sustained attack by semantic externalists in the 1970's and 1980's. One central objection was that construing proper names and natural kind terms as semantically equivalent to definite descriptions conflicts with our semantic intuitions about *de re* modal claims like 'Aristotle could have died as a child' or 'Lemons might be blue' (Kripke 1980; Putnam 1970). A further objection was that simple descriptivism underestimates the scope for ignorance and error on the part of competent speakers: one can be competent with the meaning of 'gold' without grasping a failsafe criterion for the correct application of that expression (Kripke 1980; Putnam 1970; Burge 1979). This fallibility, moreover, is crucial to the stability of meaning through open-ended inquiry and debate (Putnam 1973) and to our ability to represent genuinely objective features of our environment (Putnam 1975; Millikan 1984; Stalnaker 2008). If such objections are sound, then simple descriptivism is extensionally inadequate: its semantic competence conditions are too demanding and its determination theory generates implausible contents.

In response, internalists developed neo-descriptivist metasemantic theories. Neo-descriptivists seek to vindicate the commonsense view that the semantic function of expressions like ‘Gödel’, ‘water’ or ‘arthritis’ is to stably represent particular objects, kinds or properties, even though we may be ignorant or mistaken about the instantiation conditions of these features. Their proposal is to loosen the connection between competence conditions and semantic content. On this approach, semantic competence may consist in criteria that determine the reference only relative to empirical facts about one’s actual environment. Competence with ‘water’, for instance, may consist in having a conditional criterion for identifying the essential nature of the reference on the basis of information about the chemical nature of the liquids in one’s actual historical environment: *if* your environment is like Twin Earth *then* water = XYZ, *if* your environment is like Earth *then* water = H_2O, and so on. But you cannot know what water is without empirical information about your actual environment. So neo-descriptivist can agree that a competent speaker may be ignorant or mistaken about the essential nature of water. Even so, neo-descriptivism holds that competence with the same meaning provides a *conditional guarantee* of sameness of reference: *if* the relevant environmental facts are the same, *then* two competent uses of ‘water’ are guaranteed to pick out the same reference (Peacocke 1992, Jackson 1998a and b). Thus neo-descriptivism vindicates the core internalist idea that both competence conditions and semantic content are determined (in part or wholly) by a speaker’s internal criteria for using an expression.

In metaethics, neo-descriptivism has been used to explain competent speakers’ ignorance and error about the instantiation conditions of normative properties. According to Frank Jackson and Philip Pettit’s ‘moral functionalism’, for instance, competence with the meaning of the moral term ‘fair’ is constituted by a specific pattern of inferential, recognitional, epistemic, predictive and motivational dispositions, which together constitute the subject’s implicit ‘folk theory’ of morality (Jackson and Pettit 1995, 22-23). The semantic content of ‘x is fair’ is the property determined by the upshot of ideal reflective equilibrium, starting from this folk theory and taking into account any empirical facts that are relevant from the point of view of that theory. In a similar spirit, Christopher Peacocke holds that competence with moral terms consists in the subject’s dispositions to reason in accord with in a set of core moral principles. The applicability conditions of moral terms – and hence the properties they pick out – must be justifiable on the basis of these a priori moral principles together with empirical facts about circumstances of evaluation (Peacocke 2004). These neo-descriptivist accounts share a common internalist structure: the competence conditions allow one to identify the precise instantiation conditions for normative properties through empirical inquiry and ideal reflection. So on this account, ordinary speakers can be ignorant or mistaken about the instantiation conditions of morally rightness, even though the upshot of ideal empirically informed reflection is guaranteed to be correct.

An important advantage of neo-descriptivism is its capacity to vindicate intuitive judgments about the semantic contents of our words and thoughts. Although externalists emphasize our fallibility about *de re* necessities, their arguments typically rely on our reflective, empirically informed judgments about which properties are picked out. Because neo-descriptivism takes reference to be determined by ideal, empirically informed reflection, the approach is immune to intuitive counterexamples. Neo-descriptivism also affords a clear explanation of why all competent speakers are guaranteed to co-refer: any two individuals who rely on the same criteria must co-refer if they share the same empirical context. Moreover, the neo-descriptivist metasemantics is perfectly general: neo-descriptivism about normative terms is part of a well-motivated uniform metasemantics for of *all* terms, including names, natural kind terms, commonsense functional kind terms, and logical operators (Peacocke 1992).

2.2 meaning stability & convergence

One central difficulty for neo-descriptivism is vindicating the stability of meaning over time and between speakers. By commonsense standards, novices and experts share the same meanings when they use terms like 'water' or 'is morally right', and there is no change in meaning when rational inquiry leads novices to become more expert. But prior to further empirical inquiry the novice's ultimate reference-fixing criterion might hinge on deference to experts' criterion, whereas the expert's criterion will not. In such cases, neo-descriptivists must deny there are stable shared meanings: novices and experts rely on different criteria, which will generate divergent verdicts about reference relative to variations in the social environment.

The problem of vindicating stability of meaning is particularly acute for normative terms, since competence with terms like 'x is right' seems consistent with disagreement about any particular application of the term or any general criterion of application. Explaining meaning stability is thus the crux of the familiar disagreement problem in metaethics: how can different speakers be competent with the same meaning if they diverge in their criteria for using a term?

The disagreement problem is endemic to internalist metasemantics: it is generated by internalists' commitment to explaining competence with the same meaning in terms of matching internal criteria of use, and then using these criteria to explain the determination of semantic content. Because traditional realists claim that normative terms always pick out the very same property, the disagreement problem for realism will arise from divergence in speakers' applicability criteria. By commonsense standards, two speakers can diverge in virtually any aspect of their understanding of which actions are right without compromising their competence with the same meaning. If their criteria of application diverge substantially, how can their internal states guarantee that they single out the very same property as the reference? A similar problem arises for expressivists who identify competence with normative terms with internal criteria for using those terms to express motivational states. An internalist determination theory might then assign an expressive semantic content on the basis of these criteria of use. Intuitively, however, competent speakers can diverge in the precise motivational dispositions they associate with normative judgments like 'x is fair' or 'x is right': one can become disaffected or cynical without eo ipso losing competence with the meaning of moral terms (Merli 2008). But on an internalist metasemantics, a shift in core motivational criteria suffices for a difference in expressivist meaning. A similar problem will face internalists who favour contextualist or relativist semantic contents. In order to share precisely the same meanings, different individuals must rely on implicit criteria for using a term that relativize normative and evaluative claims to the same parameters in the same ways. But given the scope of normative disagreement, it's not obvious that there is any such convergence among competent speakers (Silk 2013).

A natural reply on behalf of internalists of all stripes is to appeal to ideal convergence in order to explain sameness of meaning. At the ideal limit of reflection, competent speakers would independently converge on precisely the same criteria for using an expression. But skeptics can argue that there is no good empirical reason to believe in ideal convergence. And appealing to ideal convergence would make sameness of meaning epistemically opaque to ordinary subjects placed in non-ideal circumstances: it may be far from obvious, for instance, whether two tokens of 'Hesperus' express the same meaning.

One might wonder why an internalist should be worried by the failure to explain meaning identity over time and between subjects. Mere similarity of meaning might suffice to explain how information is normally preserved in communication and memory.

However, meaning identity is crucial to explaining logical relations. In order for two claims to stand in a relation of direct logical contradiction (e.g. 'Hesperus is bright' vs. 'Hesperus is not bright'), the contents of the non-logical expressions must be strictly identical. Moreover, mere sameness of semantic content does not suffice for logical relations: the sameness must be guaranteed by competence with the same meaning. Consider the contrast between 'Hesperus = Hesperus' and 'Hesperus = Phosphorus'. The first claim is logically guaranteed to be true in virtue of meaning of '=' and the fact that both tokens of 'Hesperus' express the same meaning (and therefore must represent the very same thing if they represent anything at all). So if you're competent with the meaning of the sentence, its truth will seem obvious and rationally incontrovertible. Not so for 'Hesperus = Phosphorus': one may firmly believe this claim, but its truth seems to depend on contingent empirical facts and hence it won't be rationally incontrovertible. Insofar as internalists wish to explain logical relations like direct contradiction, entailment, and trivial identity of content over time and between subjects, they must offer an account of meaning identity, not just similarity of meaning. One possible response on behalf of the internalist is to deny such logical relations hold over time and between speakers using normative terms. For an internalist-friendly explanation of the surface phenomena of disagreement between subjects that does not posit such logical relations, see (Plunkett and Sundell 2013).

3. Externalism

Metasemantic externalism is simply the negation of internalism:

> Externalism: It's not the case that both semantic competence and semantic contents are determined exclusively by physical factors inside the individual or psychological resources available to the individual.

We'll focus first on externalist accounts of content determination, which have played a prominent role in defending traditional realism in both the philosophy of science and metaethics. We'll then consider externalist approaches to competence conditions.

3.1 causal theories of reference

A central theoretical advantage of externalism is its ability to explain the stability of reference despite variation in understanding over time and between subjects. In the case of names and natural kind terms, many externalists have suggested that content is partly determined by causal-historical or nomic relations linking subjects' representational states to particular objects, kinds or properties in their environment. It's important to distinguish such causal externalist theories from neo-descriptivist theories that include a subject's conception of causal roles as part of her core criteria for applying certain terms – a position known as 'causal descriptivism' (Kroon 1987). Whereas causal descriptivists take the ultimate arbiter of reference determination to be the subjects' *conception* of the relevant causal role, causal externalists take the ultimate arbiter of reference determination to be a *causal-explanatory theory* of the subject's linguistic or conceptual practices. For instance, Michael Devitt defends his causal-historical account of reference determination as meeting general desiderata on empirical theorizing in linguistics (Devitt 1981, 1991), and Ruth Millikan defends her teleosemantic account as fitting within the explanatory paradigm for theorizing about biological systems honed by natural selection (Millikan 1984). On such causal externalist accounts, an individual's current criteria for applying a term play no decisive role in settling its reference.

In metaethics, Richard Boyd's 'causal regulation' theory is the most fully developed and influential version of the causal externalist approach (Boyd 1988; see also Brink 1989; Railton 1986; Sturgeon 1985). The reference of both scientific and moral predicates, Boyd argues, depends on a causal feedback relation of 'accommodation' between a system of representations and a system of homeostatic property clusters in the world. This account allows for variabililty in speakers' criteria for applying a term without risk of changing the reference: an external causal relation can still causally 'lock' subjects' use of a word onto the same property despite differences in their internal criteria.

In the philosophy of mind and language, causal externalism has been criticized as: (i) too vague to single out a determinate reference, and (ii) singling out the intuitively wrong reference in many cases (for overviews, see (Loewer 1999; Neander 2006)). In metaethics, these general problems for causal externalism seem especially damaging in the case of normative terms, which are less tolerant of indeterminacy and less beholden to causal-explanatory considerations than scientific terms (Schroeter and Schroeter 2013).

Terry Horgan and Mark Timmons press a different objection to causal externalism: they contend that Putnam's Twin Earth argument in favor of externalism for natural kind terms like 'water' does not generalize to normative terms like 'good' (Horgan and Timmons 1992). Horgan and Timmons argue that, while we have reason to accept that causal factors affect the reference of natural kind terms like 'water', we have no such reason to accept that causal factors affect the reference of moral terms. However, the MTE argument is controversial. Janice Dowell argues that the MTE argument does not parallel Putnam's original Twin Earth argument, and that the sort of intuitions elicited by MTE are not relevant to the explanatory project of causal externalist metasemantics. She concludes MTE has no probative force against causal externalism (Dowell 2015).

3.2 reference magnets

A second type of externalist account of reference-fixing appeals to metaphysical facts about the referential candidates themselves, rather than causal relations linking representational states to referential candidates. Many metaphysicians hold that some objects, kinds, and properties are objectively more *natural* or *fundamental* than others. David Lewis suggested objective metaphysical naturalness makes a feature a better referential candidate: a property's naturalness makes it a "reference magnet" (Lewis 1983).

Lewis originally proposed this metaphysical constraint on reference determination as a way of avoiding "Putnam's paradox". Hilary Putnam argued that Lewis's global neo-descriptivism together with his metaphysical realism about referential candidates leads to radical indeterminacy of reference. In response, Lewis added a further, mind-independent metaphysical constraint on reference determination (Lewis 1984). On Lewis's account, the correct semantic interpretation of an individual's words and thoughts is determined by the total assignment of semantic contents that provides the best balance between two factors:

(i) *fit*: The assignment construes the subject's overall pattern of practical and cognitive dispositions as *approximately satisfying norms of ideal practical and theoretical rationality*, and

(ii) *naturalness*: The assignment construes the subject's words and thoughts as picking out a set of objects, kinds and properties that, considered as a group, are *overall more natural* than competing sets of referential candidates.

The problem with Lewis's earlier account was that fit alone could not single out a determinate reference.

We'd like to emphasize four points about the role of reference magnetism in a determination theory. First, the metaphysical facts about naturalness must be *entirely independent* of the subject's understanding, if they're to solve Putnam's paradox. The naturalness constraint comes into play only after we take into account the subject's ideally reflective and empirically informed dispositions to make judgments about the nature and extension of her own words. So the relative naturalness of a property is not constrained by the subject's intuitions about plausible referential candidates for a given domain. Second, the naturalness constraint is part of a *holistic* approach to reference determination. Causal theories like Boyd's link individual representations (or clusters of representations) to referential candidates on a case-by-case basis. In contrast, Lewis's Global Descriptivism provides a holistic constraint on the semantic interpretation of the entire representational system considered as a whole, which allows for interpretive trade-offs that generate semantic indeterminacy. The naturalness constraint is designed to counteract the effects of this holism. Third, any plausible interpretation must include both natural and unnatural properties in the total semantic assignment. On Lewis's account, your use of words like 'grue' or 'groovy' can pick out highly unnatural properties, provided that this interpretation is part of a total semantic assignment that maximizes naturalness overall. Fourth, it follows that there must be two sorts of ordinal rankings of naturalness on Lewis's account: (1) ranking naturalness of individual referential candidates (such as objects, properties, kinds) and (2) ranking the relative naturalness of sets of such referential candidates (which include both highly natural and highly unnatural referential candidates). Lewis himself only provides a toy example of such rankings, so objective ranking remains an outstanding issue for the approach. For a discussion of the structure and commitments of Lewis's account, see (Williams 2015).

Recently, some metaethicists have suggested that reference magnetism can be used to defend traditional context-invariant normative realism (van Roojen 2006; Dunaway and McPherson 2014). If moral properties are highly metaphysically natural (or fundamental), then they would be good referential candidates on a Lewisian approach. So a Lewisian metasemantics might be able to vindicate co-reference despite disagreement at the ideal limit of reflection, which would neutralize the disagreement objection.

However, the metasemantic story behind this approach has yet to be fully elaborated. There are general worries about whether a reference magnet approach can secure a determinate and plausible referential assignments for ordinary descriptive vocabulary (e.g. Williams 2007; Sundell 2012). In the case of metaethics, moreover, the problem is exacerbated by the fact that normative terms do not seem to be easily ranked on a single uniform scale of naturalness with causal-explanatory properties. For instance, it's not obvious that there's a principled way of ordinally ranking properties and sets of properties for 'naturalness' that would privilege a justificatory property over a sociological property as the reference of 'morally right' (Schroeter and Schroeter 2013).

3.3 externalist competence conditions

Let's turn to externalist approaches to competence conditions. According to internalism, two speakers cannot share the same meaning unless they share precisely matching internal criteria of use. Externalists deny this claim. Indeed, most externalists hold that a precise match in criteria is *neither necessary nor sufficient* for sameness of meaning or concept. Tyler Burge's anti-individualism is a version of this position:

> Anti-individualism: Semantic competence doesn't depend exclusively on physical factors in the individual or psychological resources cognitively available to the individual at a time (cf. Burge 2007, 153).

A theory of competence, of course, must go beyond this purely negative thesis: it must explain how external factors combine with internal factors to ground competence with particular meanings.

A promising approach to competence externalism is to treat *inter-cognitive relations* as necessary for competence with the same meaning:

> Relational competence: Two token cognitive states express the same meaning only if those tokens are linked by a specific causal-historical relation, R.

Relational theories of competence disagree about the nature of R. On a Kripkean "causal chain" account of names, for instance, relation R consists in being actually linked by co-referential intentions: two speakers are competent with the same linguistic meaning only if they are connected via a continuous chain of linked co-referential intentions (Kripke 1980; Devitt 1981). In contrast, Ruth Millikan's teleosemantic theory takes two uses of a term to express the same meaning only if they are connected by a certain naturally selected for 'copying' relation (Millikan 1984). Tyler Burge holds that sameness of conceptual content requires anaphora-like semantic memory relations linking token elements of thought within an individual and similar 'content preserving' causal links between individuals (Burge 2007). Philosophical proponents of 'mental files' like John Perry and Sam Cumming have posited causal coordinating relations linking individuals' mental files to those of others in their linguistic community (Perry 2001; Cumming 2013). In the philosophy of language, some theorists have argued for relationally individuated syntactic units of interpretation (Kaplan 1990; Fiengo and May 2006), and others posit relational semantic rules linking token expressions (Fine 2007; Pinillos 2011).

One important advantage of a relational constraint on competence is that it allows for more variation in competent subjects' substantive understanding than is possible on internalist accounts. Causal-historical relations between token representational states are supposed to carry some of the burden of securing semantic competence and sameness of semantic content. As a consequence, a relational account may be better placed to vindicate commonsense epistemic commitments about open-ended inquiry and debate: no particular criterion is required for logical relations among token uses of a term.

However, such variability in understanding raises a challenge: if internal criteria associated with R-linked token states are highly variable, what ensures that all R-linked tokens have precisely the same semantic content? One natural response on behalf of a relational theorist is to take the *default unit for semantic interpretation* to be the R-linked states considered as a group, rather than a token state considered in isolation (Schroeter 2012). So the default assumption would be that all R-linked uses of 'Aristotle' should be assigned the same semantic content – whether it's the shipping magnate, or the philosopher, or a descriptive content, or nothing at all. Mutatis mutandis for 'x is right': the default presumption is that all R-linked tokens should be assigned the same semantic content. For an application of this relational approach to competence with normative concepts, see (Schroeter and Schroeter 2014). (Blackburn 1991, 4-11; 1998, 59-68) also assumes social units for semantic interpretation.

One counterintuitive consequence of relational models of competence is that subjects who are not connected by R will not share the same meaning. So a perfect qualitative duplicate of you living on a perfect duplicate planet on the other side of the galaxy (Duplicate Earth) would not be competent with the same meanings that you associate with 'water' or 'right'. In response, relational theorists have explained the theoretical advantages of relationally individuated competence conditions that in their view outweigh this counterintuitive consequence (cf. Millikan 1984; Burge 2007). And the relational theorist can point out that non-relational internalist accounts also have highly counterintuitive consequences, if they make meaning unstable within rational inquiry and debate (Schroeter and Schroeter 2016).

A second challenge to relational theories is to explain how a shift in meaning is possible. If being R-related is sufficient for competence with a given meaning, then it must also be sufficient for sameness of content. But that seems implausible. For instance, our current use of 'Madagascar' may be R-related to early Malay and Arabic uses of the name which refer to a part of the African mainland, but clearly there has been a shift in reference between then and now (Evans 1973). To allow for shifts in content, a relational theorist has a number of options. One is to offer an account of R that explains how new meanings get initiated – e.g. by new implicit 'baptisms' (Donnellan 1974). A second option is to accept that contents are always preserved by R-relations, but hold that new contents get layered on top of old ones: so token states within an R-linked tradition become multiply ambiguous (Devitt 1981; Millikan 2000). A third option is to appeal to further constraints on sameness of meaning that can defeat the default presumption that R individuates the unit for semantic interpretation (Schroeter and Schroeter 2014). A fourth option is to deny that R can define a genuine meaning identity relation (i.e. reflexive, symmetric and transitive): it can only define ad hoc local relations of guaranteed of sameness of content (Fine 2007; Pinillos 2011).

4. The metasemantic construal of expressivism

One important new development in metaethics is the metasemantic construal of expressivism. Expressivism has traditionally been characterized as a theory at the level of semantics: the semantic function of normative predicates in sentences like 'x is wrong' is not to attribute a property to x, but to conventionally express the speaker's conative attitude towards x. This position generates the Frege-Geach problem: how can these expressive contents contribute to the content of complex sentences like 'Jane believes that surfing isn't wrong'? In effect, the Frege-Geach problem is to explain the role played by the expressive contents of subsentential expressions in a compositional semantic theory for a language. Given that standard compositional semantics is truth-conditional, it seems that expressivists must rebuild compositional semantics from the ground up to explain how expressive contents contribute to the contents of complex sentences. (For influential recent versions of the Frege-Geach challenge see (Zangwill 1992; van Roojen 1996 ; Unwin 1999; Schroeder 2008); for expressivist responses that reinterpret the function of logical operators, see (Blackburn 1988; Gibbard 2003)).

In recent years, some theorists have argued that expressivists can accept a standard compositional semantics that uses a possible worlds framework to specify the truth-conditions of normative sentences. To understand this proposal, it's important to bear in mind that for the purposes of compositional semantics "possible worlds" are just set-theoretic constructs whose structure is suited to modeling the contribution of subsentential expressions to the content of whole sentences. What the elements of these structures represent, if anything, is irrelevant to their role in compositional semantics. On this approach, compositional semantics functions as an autonomous domain of linguistic theorizing, with its own proprietary theoretical and empirical constraints (cf. Yalcin 2014; Glanzberg 2014). Roughly, compositional semantics tells us how the language faculty delivers formal constraints on which thought contents can be literally expressed by uses of natural language expressions. Expressivists then have an easy answer to the Frege-Geach embedding problem: they can accept whichever compositional semantics is supported by empirical linguistics.

If we reserve the term 'semantics' for formal compositional semantics and we use the term 'semantic value' for the abstract objects posited by such a theory, then different metaethical positions might be thought of as disagreeing at the level of metasemantics. Expressivists and representationalists can accept that the basic compositional structure of language is perspicuously modeled by a particular compositional theory. Their disagreement is over *what makes it the case* that normative expressions have certain semantic values within that compositional semantic theory (Chrisman 2015; Ridge 2014; Silk 2015; Pérez Carballo 2015). For closely related suggestions, see (Stojanovic 2012; Yalcin 2014)

However, the philosophical difficulties for expressivism haven't disappeared, they've just been relocated. The central questions in metaethics are not about formal semantics, but about the nature of normative thought and the use of normative language to communicate those thoughts. The fact that normative language conforms to standard compositional semantics means that the thoughts expressed by normative language must conform to the entailment, consistency, and inconsistency relations posited by the correct compositional theory. This metasemantic requirement is a non-trivial *hermeneutical* constraint on acceptable theories of which attitudes are literally expressed by normative language (Pérez Carballo 2014). The challenge for expressivists is to identify non-cognitive attitudes expressed by normative terms (in simple predications, in complex embedded contexts, in non-indicative sentences, in non-assertoric speech acts, etc.), which reflect the entailment and consistency relations posited by the compositional semantics for those sentences. In other words, expressivists must show how the semantic relations posited by the compositional theory are reflected in relations among the psychological states that they take to be expressed. This metasemantic challenge seems less daunting if one can invoke entailment and consistency relations among the *representational contents* of thoughts, rather than abstracting them from relations among the *attitudinal aspect* of thought. See (Ridge 2014) and (Silk 2015) for two recent proposals for how to meet this metasemantic challenge to attitude expressivism, and (Chrisman 2015) for an inferential role expressivism in the spirit of Robert Brandom's inferential role semantics (Brandom 1994).

Expressivists also face a challenge in explaining logical relations among normative sentences. What makes your claim 'Lying is wrong' logically inconsistent with my claim 'Lying is not wrong', if the contents expressed are simply divergent conative states towards lying? This problem is often assimilated to the Frege-Geach problem. But the semantic entailment captured by compositional semantics isn't the same as strict logical entailment: 'x is a bachelor' semantically entails 'x is unmarried', but there is no logical relation among the claims expressed. Strict logical relations are needed to explain the nature of deductive arguments across all domains. In response, (Baker and Woods 2015) argue that expressivists and representationalist alike should appeal to the syntactic structure of sentences to explain logical relations. On standard accounts of philosophical logic, after all, logical relations are purely formal: 'x is an unmarried man' logically entails 'x is a man' because it holds true regardless of how one interprets the non-logical expressions. If logical relations are determined by formal syntactic relations, then any acceptable metasemantic theory must assign the same content to the same syntactic unit. Thus the expressivist need not locate logical relations in the clashes among attitudes expressed. The availability of this formal account of logic, of course, does not guarantee that there is a plausible expressivist interpretation of the thought contents literally communicated by normative sentences. Moreover, as we noted in §3.3, such an account will need to say something about apparent shifts in the content of syntactically individuated words.

The general lesson of this recent work is that metaethicists of all stripes can help themselves to independently justified theories of the syntactic structure and combinatorial semantics of normative language. This move dissolves some traditional worries for metaethical expressivism, but structurally similar concerns may arise once again at the level of the thought contents expressed.

Acknowledgments: Thanks to the editors and James Fritz for very helpful advice and feedback.

References

Baker, D., and J. Woods. 2015. How Expressivists Can and Should Explain Inconsistency. *Ethics* 125:391–424.

Björnsson, G. 2012. Do 'Objectivist' Features of Moral Discourse and Thinking Support Moral Objectivism? *Journal of Ethics* 16:367–393.

Blackburn, S. 1991. Just Causes. *Philosophical Studies* 61:3–17.

———. 1998. *Ruling Passions*. Oxford: Oxford University Press.

Boyd, R. N. 1988. How to be a Moral Realist. In *Essays on Moral Realism*, edited by G. Sayre-McCord. Ithaca: Cornell University Press.

Brandom, R. 1994. *Making It Explicit: Reasoning, Representing and Discursive Commitment*. Cambridge, MA: Harvard University Press.

Brink, D. O. 1989. *Moral Realism and the Foundations of Ethics*. New York: Cambridge University Press.

Burge, T. 1979. Individualism and the Mental. *Midwest Studies in Philosophy* 4:73–121.

———. 2007. *Foundations of Mind*. Oxford: Oxford University Press.

Burgess, A. and B. Sherman (eds.). 2014. *Metasemantics: New Essays on the Foundations of Meaning*. Oxford: Oxford University Press.

Chrisman, M. 2015. *The Meaning of 'Ought': Beyond Descriptivism and Expressivism in Metaethics*. Oxford: Oxford University Press.

Cumming, S. 2013. From Coordination to Content. *Philosophers' Imprint* 13 (4):1–17.

Devitt, M. 1981. *Designation*. New York: Columbia University Press.

———. 1991. *Realism and Truth*. 2nd ed. Princeton: Princeton University Press.

Donnellan, K. S. 1974. Speaking of Nothing. *Philosophical Review* 83:1–31.

Dowell, J. L. 2015. The Metaethical Insignificance of Moral Twin Earth. *Oxford Studies in Metaethics* 11.

Dunaway, B., and T. Mcpherson. 2014. Reference Magnetism as a Solution to the Moral Twin Earth Problem.

Evans, G. 1973. The Causal Theory of Names. *Proceedings of the Aristotelian Society* Supp. 47:187–208.

Fiengo, R., and R. May. 2006. *De Lingua Belief*. Cambridge, MA: MIT Press.

Fine, K. 2007. *Semantic Relationalism*. Oxford: Blackwell.

Glanzberg, M. 2014. Explanation and Partiality in Semantic Theory. In Burgess and Sherman. 2014.

Horgan, T., and M. Timmons. 1992a. Troubles for New Wave Moral Semantics: The "Open Question Argument" Revived. *Philosophical Papers* 21:153–175.

Jackson, F., and P. Pettit. 1995. Moral Functionalism and Moral Motivation. *Philosophical Quarterly* 45:20–40.

Kaplan, D. 1990. Words. *Proceedings of the Aristotelian Society* supp. 64:93-119.

Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kroon, F. 1987. Causal Descriptivism. *Australasian Journal of Philosophy* 65:1–17.

Lewis, D. 1983. New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61:343–377.

———. 1984. Putnam's Paradox. *Australasian Journal of Philosophy* 62: 221–236.

Loewer, B. 1999. A Guide to Naturalizing Semantics. In *A Companion to the Philosophy of Language*, edited by B. Hale and C. Wright. Oxford: Blackwell.

Merli, D. 2008. Expressivism and the Limits of Moral Disagreement. *The Journal of Ethics* 12:25-55.

Millikan, R. G. 1984. *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.

———. 2000. *On Clear and Confused Ideas*. Cambridge: Cambridge University Press.

Neander, K. 2006. Naturalistic Theories of Reference. In *The Blackwell Guide to the Philosophy of Language*, edited by M. Devitt and R. Hanley. Oxford: Blackwell.

Peacocke, C. 1992. *A Study of Concepts*. Cambridge, MA: MIT Press.

———. 2004. Moral Rationalism. *Journal of Philosophy* 101:499–526.

Pérez Carballo, A. 2015. Semantic Hermeneutics. In Burgess and Sherman. 2014.

Perry, J. 2001. *Reference and Reflexivity*. Palo Alto, CA: CSLI Publications.

Pietroski, P. 2010. Concepts, Meanings, and Truth: First Nature, Second Nature, and Hard Work. *Mind and Language* 25:247–278.

Pinillos, N. Á. 2011. Coreference and Meaning. *Philosophical Studies* 154:301–324.

Plunkett, D., and T. Sundell. 2013. Disagreement and the Semantics of Normative and Evaluative Terms. *Philosophers' Imprint* 13 (23):1–37.

Putnam, H. 1970. Is Semantics Possible? In *Language, Belief and Metaphysics*, edited by H. E. Kiefer and M. K. Munitz. New York: SUNY Press.

———. 1973. Explanation and Reference. In *Conceptual Change*, edited by G. Pearce and P. Maynard: Dordrecht-Reidel.

———. 1975. The Meaning of 'Meaning'. *Minnesota Studies in the Philosophy of Science* 7:131–193.

Quine, W. V. O. 1951. Two Dogmas of Empiricism. In *From a Logical Point of View*. Cambridge MA: Harvard University Press.

Railton, P. 1986. Moral Realism. *Philosophical Review* 95:163–207.

Ridge, M. 2014. *Impassioned Belief*. Oxford: Oxford University Press.

Schroeder, M. 2008. *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.

Schroeter, L. 2012. Bootstrapping our Way to Samesaying. *Synthese* 189:177–197.

Schroeter, L., and F. Schroeter. 2013. Normative realism: co-reference without convergence? *Philosophers' Imprint* 13 (13):1–24.

———. 2014. Normative Concepts: A Connectedness Model. *Philosophers' Imprint* 14 (25):1-26.

———. 2016. Semantic Deference vs Semantic Coordination. *American Philosphical Quarterly* 53.

Silk, A. 2013. Truth Conditions and the Meanings of Ethical Terms. *Oxford Studies in Metaethics* 8.

———. 2015. How to be an Ethical Expressivist. *Philosophy and Phenomenological Research* 91:47–81.

Stalnaker, R. 2008. *Our Knowledge of the Internal World*. Oxford: Clarendon Press.

Stojanovic, I. 2012. On Value-Attributions: Semantics and Beyond. *Southern Journal of Philosophy* 50:621–638.

Sturgeon, N. 1985. Moral Explanation. In *Morality, Reason, and Truth: New Essays on the Foundations of Ethics*, edited by D. Copp and D. Zimmerman. Totowa NJ: Rowman & Allanheld.

Sundell, T. 2012. Disagreement, Error, and an Alternative to Reference Magnetism. *Australasian Journal of Philosophy* 90:743–759.

Unwin, N. 1999. Quasi-Realism, Negation, and the Frege-Geach Problem. *Philosophical Quarterly* 49:337–352.

Van Roojen, M. 1996 Expressivism and Irrationality. *Philosophical Review* 105:311–335.

———. 2006. Knowing Enough to Disagree: A New Response to the Moral Twin Earth Argument. *Oxford Studies in Metaethics* 1.

Väyrynen, P. 2013. *The Lewd, the Rude, and the Nasty*. Oxford: Oxford University Press.

Williams, J. R. G. 2007. Eligibility and Inscrutability. *Philosophical Review* 116:361–399.

———. 2015. Lewis on Reference and Eligibility. In *A Companion to David Lewis*, edited by B. Loewer and J. Shaffer. Oxford: Wiley Blackwell.

Yalcin, S. 2014. Semantics and Metasemantics in the Context of Generative Grammar. In Burgess and Sherman. 2014.

Zangwill, N. 1992. Moral Modus Ponens. *Ratio* 2:177–193.