

EXTERNALISM AND CONCEPTUAL ANALYSIS

Christopher A. Vogel

Vogel, C.A. (2018) Externalism and Conceptual Analysis.
Metaphilosophy, 49(5): pp. 730–65

(please cite original source: <https://doi.org/10.1111/meta.12339>)

Abstract

The method of Conceptual Analysis makes use of natural language speaker intuitions about the meaning of expressions, and relies on an externalist assumption about meanings—namely that they can be given in terms of referential relations and truth. I argue this widely used methodology is troubled, because the assumed externalist hypothesis about natural language meanings is beset with trenchant obstacles in explaining linguistic phenomena. I argue that the use of Conceptual Analysis in metaphysical investigation inherits the difficulties for the externalist hypothesis generally assumed by metaphysical inquiry. I explore two cases of such investigation that serve to exemplify the breadth of topics susceptible to troubles with the externalist hypothesis regarding natural language meanings: the extended mind hypothesis and the metaphysics of causation.

keywords; internalism, ontology, realism, polysemy, truth conditions, causation

1 Introduction

Metaphysical investigation traditionally proceeds by way of linguistic meaning. Conceptual Analysis plays a central role in this tradition, as the primary investigatory tool in addressing metaphysical questions. This familiar form of argument begins with a (fictional) case description. Readers

are then invited to consider their judgments regarding some element of the described case. which if then proffered as evidence for (or against) some metaphysical claim. The pervasiveness of this philosophical methodology is difficult to overstate. Debates about mental content (Burge, 1975, 1979), the ontology of persons (Shoemaker, 1984), time (Prior, 1967), identity (Black, 1952), modality (Plantinga, 2003), rationality (Williams, 1979), moral theory (Foot, 1967), and many others require that natural language intuitions play a profound evidential role in settling ontological disagreements.

These various vignettes, and the intuitive judgments of their audiences, are used to justify proclamations about ontology. The fertility of this tool is grounded in the underlying methodological assumptions that render its use coherent. Because Conceptual Analysis leverages the judgments of natural language speakers, claims about the nature of linguistic meaning are among these assumptions. In particular, Conceptual Analysis relies on the following *externalist hypothesis* about natural language semantics:

- (\mathcal{E}) For any expression e (in a given language L), the meaning of e determines e 's truth-conditions.

This core Fregean thought, when combined with the assumption that the judgments solicited from natural language speakers in deploying Conceptual Analysis are indicative of the world's ontology, entails that what natural language speakers understand in understanding the meanings of natural language expressions are their truth-conditions.¹ I'll argue that this contention is deeply troubled. If my arguments are successful the difficulties that plague the externalist hypothesis (E) undermine the use of Conceptual Analysis in addressing metaphysical claims, and thereby undermine the motivations for entertaining many metaphysical questions and the claims made to address them.

I begin by outlining the general critique of (E) as it applies to natural languages. The primary criticism I'll press is that natural languages exhibit a thorough-going *lexical flexibility* that cannot be explained by the rigid models available to a proponent of (E). I then show that Conceptual Analysis

¹I use 'truth condition' and its various forms as shorthand for the more cumbersome 'truth or satisfaction condition'. The more cumbersome phrase marks a distinction between the properties attributable to sentences and subsentential expressions. Sentences can be true or false, while their constituents cannot. I trust that this choice in favor of readability over precision will not be misleading.

can only justifiably be used to address questions about ontology if (E) is true for the language used in such analysis. I conclude by applying this finding to two case studies of metaphysical inquiry, to illustrate both the pervasiveness of this methodology, and the impact of the troubles noted here on its use. I'll both illustrate that the hypothesis proffered by Clark & Chalmers (1998) and the ongoing debates about the nature of causation in the wake of Lewis (1973a), both motivated by the use of Conceptual Analysis, are unmotivated—at least in the absence of a reply to the trenchant difficulties facing (E).

2 Against Externalism for Natural Language

There are three types of arguments against accepting the externalist thesis (E) for natural languages: acquisition arguments, ontological arguments, and data-driven arguments. I'll suggest briefly here the sources of the first two classes or arguments, before drawing considerably more attention the last class of arguments focused on the *flexibility* of natural language expressions. The purpose of mentioning these two sets of difficulties for externalism here is to show that the arguments from lexical flexibility discussed in the remainder of the section are not the sole source of trouble for externalist theses like (E)(cf. Vogel, 2016). Indeed, I do not take any one class of these arguments to be decisive. My goal is to draw attention to and call into question the plausibility of claims assumed by metaphysical methodology.

The first set of arguments against the externalist hypothesis for natural language draws on facts about the pattern of natural language acquisition in humans. These arguments highlight the need for an adequate account of language acquisition, namely the cross-linguistics data indicating that, when placed in a linguistic environment, nearly all children acquire a natural language by the age of four. Accounting for these acquisitional facts places limits on plausible theories regarding the nature of linguistic objects, and countenances the implausibility of externalist construals of natural languages as mind external objects (Chomsky, 1965, Ch. 1). To the degree that the externalist hypothesis (E) requires mind-external linguistic objects, these acquisition arguments tell against accepting (E).

A second set of arguments against the externalist hypothesis applied to natural languages point to the ontological requirements of adopting an externalist semantics. An externalist semantics asserts that expressions in a nat-

ural language have meanings that are relations, of a particular sort, between words and worldly objects. If meanings are relations, then the externalist is obliged to provide a viable account of the relevant *relata* in those meaning-bearing relations. If no plausible account for the ontological character of (either of) these *relata* is forthcoming, this speaks against the externalist account saddled with such a requirement. Put another way, if natural language meanings are relations between words and mind-external objects, then there must be an ontologically palatable conception of a *word*, such that objects can be related to them. The force of the ontological worry is that we have no plausible account on offer, and no obvious path to pursue in providing an acceptable proposal about the nature of words.

Of course, the most promising method here would appeal to the sonic features of human vocalizations, characterizing the nature of words and differentiating them by the noises we produce in linguistic communication. However, a difficulty with this view, familiar to phonologists, is that appealing to the physical properties of the sound waves produced by competent speakers of a language fail to categorize noises into *sui generis* kinds—at least not into kinds that correspond to words (Bromberger & Halle, 1995). That is, naturalistic inquiry into the underlying properties of words suggests that, as a viable subject of naturalistic inquiry, there are no (mind-independent) things *words*. Thus, if there is no natural kind ‘dog’ then there is no thing that can stand in relation to (all of) the (possible) dogs, as required by the externalist hypothesis.

The final set of arguments against an externalist semantics, which I’ll spend the balance of this section presenting, are data-driven arguments. They highlight the discord between 1) the pattern of meaning assignments competent speakers of the language give to expressions and 2) the truth-conditional properties attributed to expressions by externalist theories of meaning. Such arguments are data-driven because they show that externalist models cannot explain the linguistic data they are meant to explain. I’ll show that natural language expressions, as exhibited by the distribution of meanings assignments given by natural language speakers, present with a kind of *flexibility* that externalist models cannot accommodate.

The expressions that most sharply tell against the externalist hypothesis for natural language are those that have conflicting ontological commitments, whereby the single object needed to satisfy the multiple, and ontologically diverse, predicates of the expression would be quite bizarre. Externalist accounts of the meanings for the offending expressions attribute ontologically

incommensurate properties to some single object, and thereby have implausible entailments. Consider the following acceptable English expressions:

- (1) Leon wrote this book and it weighs five pounds.
- (2) The Los Angeles Times is printed and it is a corporation.

A competent speaker of English could felicitously use (1) in discussing Leon's loquaciousness, and likewise use (2) to justify a proposal about the relationship between company-types (e.g. conglomerates, corporations, partnerships, etc.) and the products they yield.

Yet for the externalist, the truth conditional meanings of expressions in (1) and (2) require metaphysically suspect objects. On one acceptable reading of (2), an externalist account of this reading would require the existence of a rather bizarre object.² On this reading the expression 'this book' seemingly refers to both a content (as an abstract object) written by Leon (in his head, possibly), and a physical, bulky tome. However, the expression 'this book' cannot refer to both of these distinct objects—rather, the expression must have some single object as its referent, in order to satisfy both predicates.³ This single referent cannot be the physical object of the text, since speakers would treat (1) as true even in a situation whereby Leon never interacted with, much less penned, the physical object. Likewise, the single referent of 'this book' cannot be the abstract content written by Leon (entirely in his mind, say). Abstract objects surely have properties, but mass is not among them. On either of these accounts of the meaning determined referent of 'this book', the sentence would always be false, which belies the manner in which speakers would treat this claim.

An externalist attempt to address these data by appeal to ambiguity, and indicating that expressions like 'book' are simply homophonous, ignores

²There is a different reading of (2) which of course would not require such a suspect object for the externalist. Rather, this reading indicates that Leon penned a lengthy manuscript, in a manner that would likely cramp his hand. However, any semantic theory must account for the *entire* distribution of meanings speakers assign to an expression, which in this case includes the one outlined above.

³An externalist might insist that the needed object is not that bizarre, but is simply a combination of the two referents discussed above—a mereological sum of the content and text. However, if this sum was indeed the referent of 'this book' then (1) would necessarily be false since that *sum* does not weigh five pounds, and is not written by Leon, even if these properties are applicable to *parts* of this combined object.

the fact that ‘book’ is only used a single time in (1). So even if ‘book’ is ambiguous between two homophonous expressions (one for contents and the other for mediums), only one of them can be solicited by the *single* use of book in (1). The more pressing worry for the externalist appeal to ambiguity is not merely that such an appeal fails to leave open the possibility that (1) is true, but that such an account fails to explain the inferences natural language speakers would make regarding (1). That is, natural language speakers would endorse the inference from (1) to either of the following:

(3) Leon wrote a book.

(4) A book weighs five pounds.

For the externalist, such inferences are modeled on entailment. Given the conjunctive nature of (1), the expressions in (3) and (4) are licensed inferences insofar as they are (more or less) the two conjuncts in (1). But if ‘book’ were ambiguous, the referent of ‘this book’ in (1) would only license *either* (3) or (4) given the masslessness of abstract contents.

Likewise, ‘the Los Angeles Times’ in (2), on an externalist semantics, must have a single referent that is both printed and a particular kind of company. Given that the anaphoric ‘it’ must derive its referent determining meaning from ‘The Los Angeles Times’, these phrases must co-refer to some single object, on an externalist semantics. Though abstract institutions can have many properties, they are not capable of being typeset and rendered to paper. And while physical instances of a publication can be put to paper, they do not have a corporate structure, even if the companies that produce them do. Yet, for (2) to be true there must be an object of this odd sort (cf. Chomsky, 1977; Pietroski, 2005; Vogel, 2016, for related examples).

But even if the externalist is willing to admit into their ontology the metaphysically bizarre objects needed to accommodate these cases of lexical flexibility, as Pietroski (2005) notes, the theory will still fail to explain the relevant data—namely that sentences like (5) are deviant:

(5) # The Los Angeles Times is a printed corporation.

Both (2) and (5) are perfectly grammatical constructions, and on an externalist semantics, truth conditionally equivalent. But (5) is aberrant in a way

that (2) is not. The problem for the externalist is that if both sentences express the very same truth-conditions, as determined by their differing meanings, they then owes us an explanation for the oddity of (5) in expressing the thought purportedly conveyed felicitously by (2).

Some natural language expressions exhibit a kind of flexible behavior that, while less decisive than those just explored, counts against adopting an externalist hypothesis for languages with such expressions. These expressions convey a constellation of *related* concepts, with distinct satisfaction conditions. Color terms demonstrates this kind of flexibility. Consider the following:

- (6) The car is purple.
- (7) The pencil is purple.
- (8) The ink is purple.
- (9) The carrot is purple.
- (10) The sky is purple.

The predicate ‘is purple’ attributes related properties in each of these uses, all of which are connected to our phenomenological experience of color. However, the kind of object predicated in each example seems to demand a different set of conditions from the color term. In (6), the car’s exterior must be (mostly) purple in order for the expression to be true. But the car’s interior can be any color. In direct contrast, (7) would be true if the exterior of the pencil was completely black, so long as the interior marking compound is properly colored. And even though (9) is true in cases where the relevant plant is mostly green, the ink in (8) might look to be completely black without undermining the truth of the sentence—since (8) is still true if one could scribe purple letters with the indicated fluid. Finally, (10) is true irrespective of the fact that the bits of matter that form the sky are entirely devoid of color.

These data points suggest that expressions like ‘purple’, though deeply related to the visual experiences had by humans, solicit concepts with conflicting satisfaction conditions. This in turn suggests that we would adopt a pluralist view about expressions like ‘purple’, whereby they can be used to express a variety of concepts PURPLE. We might be tempted here by the externalist inclination to suppose that these uses are just cases of homophony,

with many different words ‘purple’ all with a single extension. At the limit, we might even hold that color terms are a kind of indexical (Rothschild & Segal, 2009).

However, to treat these expressions, which are all deeply related to one another by way of human visual experience, as cases of homophony/ambiguity fails to respect the phenomena. The paradigm cases of homophony treat distinct, unrelated, concepts as expressible with different words having the same phonological form. Consider the following:

(11) The ink is in the pens.

(12) The pigs are in the pens.

The uses of ‘pens’ in these sentences express two distinct and unrelated concepts: PEN_w which relates to the instruments used in writing, and PEN_c which relates to the structures used to contain agrarian animals. These two concepts are clearly unrelated, having come to be associated with the same phonological form by historical accident. The externalist supposition here is that we have two distinct words ‘ pen_w ’/‘ pen_c ’ each of which is related to a single concept/extension. While such a claim seems plausible for unrelated uses of ‘pen’, the same simply cannot be said for deeply related uses of ‘purple’.

Importantly accepting a pluralist account of the meaning of color terms undermines the externalist hypothesis (E) since the meanings of such terms will not determine their extensions. Even if the various concepts of (say) PURPLE proposed by the pluralist reply had determinate extensions, the meaning of ‘purple’ will not determine its truth conditions, at least not without the surfeit of extra-linguistic contextual information required to arrive at the concept intended by the use of the expression. Thus, lexically flexible expressions are difficult to account for under an externalist semantics, and given the that such flexibility is rampant in natural language, they pose serious obstacles for the externalist hypothesis (E).

The arguments suggested here against (E) are not intended to be exhaustive, or even decisive. The project of dismantling the longstanding tradition of externalism in philosophy and linguistics surely cannot be completed in such a short space.⁴ The purpose of presenting these arguments here is

⁴For extended attempts at this project see Pietroski (2005, 2008, 2010, 2018) and Hinzen (2006, 2007).

merely to highlight that the externalist hypothesis (E) is contentious, and troubled, to a greater degree than philosophers tend to appreciate. Moreover, this hypothesis is central to traditional metaphysical methodology, and indeed required if we are to draw metaphysical conclusions from the use of methods like Conceptual Analysis. If the theoretical foundations of Conceptual Analysis are indeed fraught with the difficulties suggested in this section, then the conclusions wrought by these methods are equally fraught. The argument here is a conditional one: if the externalist hypothesis (E) is false, then Conceptual Analysis (at least when such analysis makes use of natural language) cannot reliably yield verdicts pertinent to ontological investigation. In the remaining sections I'll argue that indeed (E) must be true if Conceptual Analysis is to be useful in justifying metaphysical claims, and this conclusion, when paired with the problems suggested above for the externalist hypothesis, undermines the motivations for many metaphysical arguments that rely on the products of Conceptual Analysis.

3 Against Conceptual Analysis

Conceptual Analysis is a pervasively used methodology for adjudicating metaphysical theses. The success of this method in settling metaphysical questions hinges on the broader externalist hypothesis (E). The problems presented for (E) bear on metaphysics writ-large, insofar as they undermine the use of Conceptual Analysis in resolving metaphysical disputes. The barriers to accepting the externalist thesis pose a dilemma for metaphysicians⁵ hoping to use Conceptual Analysis as means of settling ontological debates, or so I argue.

I'll do this by way of example, highlighting what I take to be the foundational mistakes in much metaphysical speculation. To start, I will indicate explicitly what I take the preceding arguments to have shown about the prevalent methodology adopted by metaphysicians. These conclusions undercut particular, recent arguments regarding the extended mind hypothesis.

⁵The main target here is a metaphysician that adopts a *realist* position about ontology, in contrast to pluralist or fictionalist positions about “metaontology” (Eklund, 2006). I take the realist position to be the predominate view adopted (implicitly) by both metaphysicians and philosophers engaged in metaphysical inquiry across philosophical domains. For a paradigmatic application and defense of this view see, respectively, Sider (2002, 2011).

The broader worry for metaphysicians, and much metaphysical speculation throughout philosophy, is that this argument generalizes, applying to any domain that relies on the judgments of natural language speakers as evidence for a proposal regarding the truth-conditional definition of any concept or term apt for use in metaphysical proclamations. The application of this same worry to debates about the metaphysics of causation highlights the generality of the problem posed here. Put more succinctly, if these arguments are correct, many classic and contemporary metaphysical debates are misguided, since they misconceive the *explanada* in their respective domains.

3.1 Externalism and Conceptual Analysis

The use of Conceptual Analysis to evaluate metaphysical claims requires an externalist semantics for natural language, in the manner discussed in §2. This externalist thesis is deeply troubled. While the arguments offered above fall well short of showing that externalism is false, in the case of natural language they pose a serious, and recalcitrant problem for any theorist committed to the idea that natural language expressions have meanings that determine their truth-conditions. If the metaphysician's use of Conceptual Analysis commits her to this externalist idea, the problems for (E) render her methodology without a firm foundation. I'll argue in this section that indeed the use of Conceptual Analysis as a tool for investigating ontological questions is only coherent if the language used to solicit intuitive judgments has an externalist semantics—that is, if (E) is true for that language. Given that the language most often used in applications of Conceptual Analysis is a natural language, the failures of (E) in capturing natural language meanings undermine the utility of Conceptual Analysis in metaphysical inquiry.

The following kind of argument is quite familiar, and if I'm correct, deeply troubled:

[ROCKS]

Billy and Suzy throw rocks at bottles. Suzy throws first, or maybe throws harder. Her rock arrives first. The bottle shatters. When Billy's rock gets to where the bottle used to be, there is nothing but flying shards of glass. . . So Suzy's throw causes the shattering. Billy's doesn't. . . In such cases, we get the right answer if we take causation to be the ancestral of [counterfactual] dependence. (Lewis, 2000, p. 184)

Lewis (2000) offers this short fictional happening as evidence regarding a particular (counterfactual dependence) theory about the nature of *causation*. The familiar method is to present a case which makes use of the metaphysical notion in question, and in light of the reader’s comprehension of the passage, leverage their intuitive judgments about the described case with regard to that notion in order to purportedly adjudicate between competing theories. Here, the case is presented to show a flaw in a simple counterfactual notion of causation, and lend support to an ancestral-counterfactual account.⁶ The particulars of this debate need not concern us just yet. But what is our concern is that this kind of familiar inquiry is taken quite seriously, in what many consider to be core questions in metaphysics, like the nature of causation (cf. Collins et al., 2004).

This form of argumentation can be characterized more formally as follows:

CONCEPTUAL ANALYSIS

1. L -speakers understand the meaning of expression e ;
 2. Theory T holds that e has truth-conditions e_{def} ;
 3. ϕ is an expression that has e as a constituent;
 4. Consider case C ;
 5. C describes a truth-maker of ϕ , according to T ;
 6. Theory T predicts that L -speakers will judge expression ϕ “True” (or “False”) of C ;
 7. L -speakers *in fact* judge expression ϕ “True” (or “False”);
- \therefore Theory T makes the right (or wrong) prediction.

So formalized, Conceptual Analysis bears little resemblance to the passage above from Lewis (2000). But indeed his usage is an instance of this form. After all, readers of the passage understand, or make use of, the expression (e) ‘cause’. Lewis offers a counterfactual account (Theory T) thereby indicating the truth-conditions for this term (e_{def}), that ROCKS (Case C) is meant to continence. The case involves a truth-maker for ‘Suzy’s throw

⁶I discuss this case at length in §5 below.

causes the shattering, and Billy's does not' (ϕ) which contains the expression 'cause' (e). This counterfactual theory makes a prediction about how readers will evaluate the truth of this sentence (ϕ). In this particular case, the account makes the wrong prediction, which counts against the simple counterfactual account (Theory T).

Lewis' prose looks dissimilar from the more formal steps above because, like most who use the method, he omits an explicit endorsement of 1 (as well as (E)). Further, he phrases 6 and 7 in terms that do not refer to speakers of a language, but simply assert that the relevant expression *is* true/false (given the case). Lewis after all baldly asserts that Billy simply didn't cause the bottle to shatter. Both of these differences can be easily explained. Lewis assumes that meanings must be related to truth conditions (Lewis, 1973a). This assumption renders an appeal to language, and speakers' judgments about sentences, redundant. If one assumes the case description to be true, and that the meanings of the expressions used to describe the case determine the way the world must be for such claims to be true, Lewis can exploit his own competence with those expressions to make claims about what is (assumed to be) true according to the case description. But of course, if such claims are accurate, Lewis' competence with English must reflect competence with calculating the truth-conditions of English sentences. This just is an endorsement of (E). If this assumption is false, then he is not justified in arriving at truth-condition judgments by way of meaning judgments.

Readers might also find this form of argument unfamiliar for different, though related reasons. Omitting explicit reference to a language, and the judgments of speakers of that language, blurs the theoretical significance of presenting counterexamples as cases against theories that define the truth-conditions for some expression. One further obfuscation is the habit of labeling this form of argument "conceptual analysis." Arguments of this form are typically worded as involving, not expressions, but *concepts*.⁷ The purported counterexample to the disputed theory about the truth-conditions of some particular *concept* is meant to show that the particular *concept* humans bring to evaluating the counterexample has truth-conditions that diverge from those offered by the theory.

⁷I use 'concept' here as, minimally, mental particulars used to think about things (as such). One might use 'concept' in some other way, as (say) *abstracta* that stand in inferential relations (Peacocke, 1992). In either case, maintaining the use of Conceptual Analysis requires adopting a view that straight forwardly maps word (meanings) to concept (contents).

However, to indicate that a case-based counterexample sheds light on a particular concept is only viable if the relationship between linguistic expressions and concepts is fairly determinate. While human understanding of language is surely mediated by the conceptual system, the relationship between language and the conceptual system must involve a direct (one-to-one) mapping between words and concepts for the form of argument above to yield straightforward verdicts about the nature of *concepts* (or their contents). This would require the proponent of Conceptual Analysis to adopt a *labeling theory* of linguistic meaning. In brief, this view holds that the meanings of linguistic expressions are just conceptual contents. Words, on this view, merely *label* concepts, making their contents expressible *via* vocalization. The truth-conditions that meanings determine, on this view, are the truth-conditions had by complete thoughts. Thus when we think true thoughts, we make use of concepts that track real-world distinctions in veridical ways.

Given the history of semantics one can confidently speculate that these assumptions trade on the relationship between truth and meaning. If both linguistic meaning and conceptual content are characterized by way of truth, then one can map the (truth-conditional) meanings of linguistic expressions onto the satisfaction conditions of (truth-tracking) concepts. This distinction between expression meanings, conceptual contents, and their (would-be) wordly satisfiers is glossed over in much of contemporary philosophy that seeks to draw metaphysical conclusions from case-based judgments. Consider the following exemplar from DeRose (1999), wherein he defends the thesis that *knowledge*—not merely ‘know’, ‘knowing’ or ‘knowledge’—is sensitive to context:

Contextualism [about knowledge], as described above, is a thesis about *knowledge attributing and denying sentences*. But, since there are other *terms with analytic ties to the concept of knowledge*, we should expect that if *contextualism about knowledge* is true, there should be corresponding shifts in the content of sentences containing those other terms. (DeRose, 1999, p. 189) (my emphasis)

DeRose marks the distinction between the “terms” that serve as constituents of sentences, the concepts to which they are “tied” that serve as the constituents of thoughts, and the (would be) extension of that concept of interest to epistemologists, namely *knowledge*. In the process of articulating these

distinctions, he likewise directly assumes that judgments about *sentences* are informative about both the concept KNOWLEDGE, and what counts as *knowledge*. This bold claim seems plausible if one assumes that the meaning of a sentence determines the truth-conditions of that sentence, as DeRose does (ibid., 187), and if the relationship between the lexicon and the conceptual system is fairly tight as implied by the above passage.

I suggested above reasons to deny that sentences have meanings that determine truth-conditions. Likewise, the labeling theory is fairly implausible (cf. Glanzberg, 2011; Pietroski, 2018). However, even if we indeed assume a labeling theory of meaning, the problems stemming from the lexical flexibility of natural language still remain. If linguistic expressions are merely labels for (composed) thoughts who are the true bearers of meaning, then these conceptual meanings are flexible in exactly the way linguistic expressions appear to be. Adopting a labeling theory of meanings shifts the problems presented in the previous section, it does not address them (Vogel, 2016). For all I've said here, one can conjecture that arguments of the form above are directly informative about the content of our concepts, but that conjecture requires some substantive defense in light of the concerns expressed in the previous section.

To put this same point in different terms, and with less pace, let me dwell here on the use of Conceptual Analysis. After all, this method of inquiry has been central to philosophical argumentation for some time, and like many traditions, is often accepted without much reflection. In utilizing Conceptual Analysis the metaphysician wants to convince us that some theory about the nature of some (natural) kind is (in)correct; e.g., causation is a relation of ancestral counterfactual dependence. She does this by offering an argument, whereby she presents a case that supports (or belies) the theory in question (about, say, causation, or the Doctrine of Double Effect, or the extended mind hypothesis). Her argument, and the description of the case is expressed *via* some language or other. But that language, whatever it might be, does not trade in concepts. The case is described in sentences that make use of expressions in that language, not concepts of the human mind. Given additional (adventurous) assumptions about the relationship between linguistic meaning and conceptual content, the method of argumentation *might* serve as a means of analyzing conceptual contents. But we should mark those assumptions when they make claims about the architecture of the human mind, and especially when they are used to make claims about what exists in the mind-external world.

Conceptual Analysis leverages the truth-conditions described in some case against the proposed truth-conditions for the expression under investigation. As a speaker of the language used to pen the example, readers are asked to entertain some case description as if it were truthful, and then asked to judge the truth-value of some statement about the described scenario. Put again more technically, a case C describes (in language L) some happening, wherein the expressions of that description are proposed to have truth-conditions. L -speakers are asked to suppose those truth-conditions are satisfied by objects in the world. The theoretical upshot is purportedly had when L -speakers are then asked whether some sentence ϕ , which contains the expression of interest e , is also made true by the same object-satisfiers of the descriptions that constitute C . The L -speaker's verdict (ostensibly) speaks as to whether the truth-conditions that are offered up for e by T are apt or not. And this in turn indicates what objects there must be in the worldly domain, as constituents of the (would-be) truth-makers of the claims in C , and the satisfiers of the truth-conditions offer by T for e .

The method of Conceptual Analysis trades on the assumption that expression meanings determine their truth-conditions, as can be made fairly clear upon examining the method. A toy example involving the expression 'chair' will make the methodology clear:

Theory T :

y is a chair *iff* y has four legs, a seat, and a back...

formally:

$\llbracket \text{is a chair} \rrbracket(y) = \top$

... *iff* $[\lambda x. \text{CHAIR}(x)](y)$

... *iff* $[\lambda x. \text{FOUR-LEGGED}(x) \ \& \ \text{SEAT}(x) \ \& \ \text{BACK}(x)](y)$

... *iff* $\text{FOUR-LEGGED}(y) \ \& \ \text{SEAT}(y) \ \& \ \text{BACK}(y)$

Case C (as described in English):

Mel is made of wood. Mel has a seat. Mel has three legs. Mel does not have four legs. Mel has a back.

$\phi =$ Mel is a chair.

English-speaker judgment:

ϕ is true (of C).

The truth-conditions expressed in the sentences that constitute C are the conjunction of the following claims:

[[Mel]] = m (an object in the real-world domain)

[[Mel is made of wood]] = $\top \dots$

\dots iff $[\lambda x. \text{MADE-OF-WOOD}(x)](m)$

\dots iff $\text{MADE-OF-WOOD}(m)$

[[Mel has a seat]] = $\top \dots$

\dots iff $[\lambda x. \text{SEAT}(x)](m)$

\dots iff $\text{SEAT}(m)$

[[Mel has three legs]] = $\top \dots$

\dots iff $[\lambda x. \text{THREE-LEGGED}(x)](m)$

\dots iff $\text{THREE-LEGGED}(m)$

[[Mel does not have four legs]] = $\top \dots$

\dots iff $\neg[\lambda x. \text{FOUR-LEGGED}(x)](m)$

\dots iff $\neg\text{FOUR-LEGGED}(m)$

[[Mel has a back]] = $\top \dots$

\dots iff $[\lambda x. \text{BACK}(x)](m)$

\dots iff $\text{BACK}(m)$

Taken in conjunction then, C is true just in case the following is true:

$\text{MADE-OF-WOOD}(m) \ \& \ \text{SEAT}(m) \ \& \ \text{THREE-LEGGED}(m) \ \&$

$\neg\text{FOUR-LEGGED}(m) \ \& \ \text{BACK}(m)$

If this expression is true, then the *satisfiers* of the five predicates are indicated therein. That is, m satisfies the truth-conditions of the related linguistic expressions in C .

According to Theory T , the truth-conditions for ‘Mel is a chair’ are:

$$\llbracket \text{Mel is a chair} \rrbracket = \top \text{ iff } [\lambda x. \text{FOUR-LEGGED}(x) \ \& \ \text{SEAT}(x) \ \& \ \text{BACK}(x)](m)$$

or simplifying:

$$\llbracket \text{Mel is a chair} \rrbracket = \top \text{ iff } \text{FOUR-LEGGED}(m) \ \& \ \text{SEAT}(m) \ \& \ \text{BACK}(m)$$

As such, Theory T predicts that English-speakers will judge ‘Mel is a chair’ to be false in C . The truth-conditions of the expressions used to describe C indicate that ‘ $\neg\text{FOUR-LEGGED}(m)$ ’ is true, rendering ‘ $\text{FOUR-LEGGED}(m)$ ’ false. Thus, that English-speakers judge ϕ to be true, holding that Mel is indeed a chair, serves as evidence against Theory T . This analysis is meant to tell us about our concept CHAIR (as distinct from the word ‘chair’), highlighting the fact that the objects in the domain that meet the satisfaction conditions of the CHAIR-concept need not have four legs. This supports the metaphysical claim that an object can be a chair even if it fails to be four-legged.

Admittedly, this is a toy example focused on a sophomoric case, with quite impoverished indications of the semantics for the English expressions involved. However, what should be clear is that the theoretical upshot of this methodology, when applied to a domain of purported philosophical interest, only gets traction on the relevant conception (e.g. CHAIR) if we assume both that the meanings of English language expressions (like the ones used in C) determine their truth-conditions, and those truth-conditions can be given in terms of conceptual contents that have real world satisfiers (e.g. m).

In reading the case in C , English-speakers understand the meanings of the expressions used therein. On the hypothesis that what said speakers understand are the truth-conditions of those expressions, this method can (straight-forwardly) inform us about the meanings of expressions like ‘chair’. On the further hypothesis that the satisfiers of those truth-conditions also serve as satisfiers of concepts (e.g. CHAIR) that have similar satisfaction-conditions, this method gives us an analysis of our human concepts. The discussion of lexical flexibility should indicate the rather bold character of these two hypotheses. In the absence of a rather substantive defense of the externalist hypothesis’ failure to account for the flexibility of natural language

expressions, such a method should not be adopted by philosophers hoping to get traction on metaphysical questions.

3.2 An Objection

One might stop at this juncture and object as follows:

But surely the use of Conceptual Analysis is not (typically) intended to shed light on the concepts that individual human minds *actually* deploy. The use of imaginative cases like ROCKS are meant to inform us about what the content of our concepts *should* be. They are intended to sharpen and clarify muddled concepts we bring to the table, in order to offer a *better* conception of the world. So whether or not English has a semantics that is externalist or not is orthogonal to the purpose of Conceptual Analysis.

Conceptual Analysis may, given certain adventurous assumption about the structure of the human mind, shed light on the content of the concepts we have. Further, the methodology may be quite useful in indicating the rather unprincipled way in which we apply them, given these assumptions. And we might further grant that a human mind with fewer unprincipled concepts might be, in some sense, a better one. The further claim the metaphysician must endorse in using Conceptual Analysis to settle debates in ontology, is that these sharpened concepts are (more) indicative of what exists.

But even if one concedes that one ought to posit in one's ontology those entities that our best inquiry requires,⁸ Conceptual Analysis fails to be instructive in this way. If metaphysical inquiry is a normative project about what expressions (or their purported conceptual contents) *should* mean, then the intuitive judgments of readers to vignettes like ROCKS serve as little justification for any particular metaphysical claim. The normative argument implies that the concepts sharpened by Conceptual Analysis are more indicative about what entities are in the worldly domain. But understanding a sentence, and passing a truth-judgment about that sentence given some case, requires that the deployment of concepts already stocked in the conceptual system, as *expressed by a natural language*—in whatever complex way that happens. Given the assumption that the meanings of sentences determine

⁸One might balk at this concession in light of the lexical flexibility exhibited by core scientific terms in the biological sciences (Vogel, under review).

their truth-conditions, Conceptual Analysis may be useful in indicating that deploying these (unsharpened) concepts yields inconsistencies regarding the truth of the constellation of claims that we comprehend by using the sentences that express these dull concepts. But it's precisely this assumption that is mistaken.

In what remains, I outline the use of this strategy, and the method of Conceptual Analysis that appropriates it, in the context of two domains: the extended mind and the nature of causation. The purpose of presenting both cases is one of contrast. The very same assumptions that might lead us astray in thinking that the boundaries of the human mind extend beyond the skull, likewise obfuscate the *explananda* in debates about the nature of causation. Noting the troubled nature of this strategy will serve to clarify the basic metaphysical question in both domains, and suggest that the vexing philosophical questions therein are either not that vexing, or at least unmotivated by considerations aroused by appeal to natural language speaker judgments. The broader conclusion is that analogous worries arise in a vast array of core metaphysical debates that likewise rely on case-based intuitions to adjudicate thesis about what there is.

4 Against Extended Minds

Since the publication of Clark & Chalmers (1998), the proposal that human minds have boundaries that extend well beyond our bodies has received considerable attention. The central proposal suggested by Clark & Chalmers (1998) is that, counter-intuitively, the human mind extends well beyond the boundaries of the human organism.

The main contention of the extended mind hypothesis is that features of the external environment constitute parts of the human mind. The introductory example given by Clark & Chalmers (1998) as a means of clarifying their thesis involves the video game *Tetris* (Pajitnov & Pokhilko, 1984). In this video game players are tasked with arranging two-dimensional shapes into a block formation. Critical to success in this task is the ability to quickly judge whether an individual piece will fit into an opening in the block formation. Because players can rotate the two dimensional pieces clockwise ninety degrees by pressing a button on the game's control pad, the player has some control over where each piece will go, and how it will fit into the block.

Assessing whether the piece can fit into a given place in the block-formation

can be accomplished in (at least) two ways: either by the player imagining the various ways in which the piece can be orientated, rotating this imagine “in her head”, or by pressing the button on the control pad to rotate the piece on the video screen, and checking the fit of the piece using her visual system. As it turns out, the latter strategy tends to be significantly faster. Clark & Chalmers (1998) indicate that in contrasting these two strategies, the function played by aspects of the mind used in imaginatively rotating the shape (in the first case), and the function played by the gaming system’s rotation algorithm (in conjunction with the player’s visual system, in the second case), are the same. Yet, only in the former case do we conclude that the procedure the player engages is completely a mental one. To Clark and Chalmers this distinction seems to be without any basis.

Their controversial claim however is not merely that, on occasion, some processes that have all the appearance of cognitive processes in fact involve organism-external objects. The claim is that key “core” components of the mind, like beliefs and desires, can be constituted by organism-external objects. If some organism-external object plays the same functional role as an organism-internal object in paradigmatic, core mental processes like belief, then (barring some other robust reason to the contrary) organism-external objects can be components of the human mind.

The primary argument presented for this view in Clark & Chalmers (1998) involves a contrast between two cases, meant to “argue that *beliefs* can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive processes” (Clark & Chalmers, 1998, p. 12). The first case that is supposed to illustrate that the nature of *belief* (of a certain sort) involves recall from memory, while the second case shows that the role played by memory in the paradigmatic case of belief can be filled by an organism-external object, i.e. a notebook. In that vein, consider:

[INGA]

Inga hears from a friend that there is an exhibition at the Museum of Modern Art, and decides to go see it. She thinks for a moment and recalls that the museum is on 53rd Street, so she walks to 53rd Street and goes into the museum. (Clark & Chalmers, 1998, p. 12)

Clark and Chalmers conclude from this case involving Inga, that she “clearly believes that the museum is on 53rd Street” and that (because of

this) *beliefs* can be stored in memory (Clark & Chalmers, 1998, p. 12). They come to this conclusion for the same (errant) reasons that most philosophers come to metaphysical conclusions based on intuitive evidence from cases of this sort—namely that the meanings of expressions determine their extensions, and thus determine what kinds of objects populate the worldly domain.

Suppose *arguendo* that we take their data point as given, namely that English-speakers would assent to the truth of ϕ on the assumption that all of the expressions in INGA are true:

ϕ = Inga believes that the Museum is on 53rd Street.

Such a data point does not permit the conclusion that *beliefs* can be stored in memory (even if we assume that meanings determine extensions). After all, Inga's belief that motivates her to head toward 53rd Street could simply be the *occurrent* belief she is consciously entertaining, relevantly divorced from her memory. To reach the conclusion that beliefs can be constituted by memory, even under the assumption that this case-driven methodology is justified, requires a slightly different case than the one they propose. Consider then:

[INGA*]

Inga hears from a friend that there is an exhibition at the Museum of Modern Art, and decides to go see it. [She stops to tie her shoes.] She thinks for a moment and recalls that the museum is on 53rd Street, so she walks to 53rd Street and goes into the museum. (Clark & Chalmers, 1998, p. 12)

Given that the expressions in INGA* are true, plus the assumption that the methodology of Conceptual Analysis is justifiable, Clark and Chalmers can reach their conclusion regarding the relationship between *memory* and *belief* if English-speakers would find the following true (of INGA*):

(ϕ^*) Prior to tying her shoes, Inga believed that the Museum is on 53rd Street.

Of course, even if we assume that English-speakers have the judgments Clark and Chalmers need, such data only justifies the claim that *beliefs* can be constituted by memory if we assume that the meaning of the English language expression 'belief' determines its extension. That is, the nature of

belief is only enlightened by such judgments if the meanings of expressions like ‘thinks’ and ‘recalls’ reflect what must be true of the world in order for speakers to judge that the relevant claim using ‘believed’ is also true. Again, the idea behind Conceptual Analysis, and the externalist assumption that underwrites the methodology, is that English-speakers’ judgments regarding ϕ^* will accord with the (purported) fact that the objects required to satisfy the truth-conditions of the sentences in INGA* will also satisfy ϕ^* . But again, this methodology only yields the metaphysical verdict that (say) *beliefs* are partially constituted by memory, if the externalist thesis (\mathcal{E}) holds. Insofar as the language used to describe INGA* is a natural language (i.e., English), the argument from lexical flexibility offers compelling reasons to think these metaphysical conclusions are unjustified.

Yet Clark and Chalmers insist that cases like INGA* offer compelling evidence about the nature of *beliefs*. In this vein, they offer up the case of OTTO’ to contrast with INGA*:

[OTTO’]

Otto suffers from Alzheimer’s disease, and like many Alzheimer’s patients, he relies on information in the environment to help structure his life. Otto carries a notebook around with him everywhere he goes. When he learns new information, he writes it down. When he needs some old information, he looks it up [in the notebook] . . . Today, Otto hears about the exhibition at the Museum of Modern Art, and decides to go see it. [He stops to tie his shoes.] He [then] consults his notebook, which says that the museum is on 53rd Street, so he walks to 53rd Street and goes into the museum.⁹ (Clark & Chalmers, 1998, pp. 12–13)

Contrasting Otto’s actions with Inga’s, Clark and Chalmers conclude “that when it comes to belief, there is nothing sacred about the skull and skin” because, much like Inga, Otto had a belief about the museum’s location prior to walking toward it (Clark & Chalmers, 1998, p. 14). Otto’s notebook, they claim, plays the same explanatory (and hence functional) role in the most plausible psychological account for his actions as Inga’s memory plays in an

⁹As with INGA, the description of Otto’s case will fail to yield the verdict Clark and Chalmers require without some additional information to distinguish occurrent beliefs from stored beliefs. It should also be noted that Clark and Chalmers do not give the cases names. These have been added for the purpose of clarity.

analogous explanation for her actions. Thus, if Inga has a non-occurrent belief about the museum's location, so too does Otto, despite his reliance on a notebook. Insofar as beliefs constitute a core aspect of the human mind, Otto's notebook must be considered a part of his mind.

This seems to imply that, according to Clark and Chalmers, English speakers would accept the following as true:

(ϕ') Prior to tying his shoes, Otto believed that the Museum is on 53rd Street.

Supposing that English-speakers indeed find ϕ' to be true (given OTTO'), such evidence does not yield the bold metaphysical conclusion about the nature of belief without assuming that the (troubled) externalist thesis (\mathcal{E}) is true.

However, Clark and Chalmers do not predict that English-speakers would assent to the truth of ϕ' given the case described in OTTO'. They indicate that the thesis being defended is not about "common usage; [their] broader point is that the notion of belief *ought* to be used so that Otto qualifies as having the belief in question" (Clark & Chalmers, 1998, p. 14). The claim here is an instance of the consideration raised in §3.2, whereby Conceptual Analysis is put to work in characterizing what a word *should* mean. The extended mind hypothesis then is a claim about how the expression 'belief' *ought* to be used, such that the extension of the term counts Otto as having a belief about the location of the museum (prior to consulting his notebook).

A charitable¹⁰ reading of their normative claim here suggests that the extended mind hypothesis regards, not the natural language expression 'belief' (or 'mind'), but is rather a proposal for a scientific term 'belief_s' (or 'mind_s'), whose meaning is the "sharpened" concept BELIEF_s (or MIND_s), whereby the term/concept has an extension that includes Otto's notebook (as described in OTTO'). Their proposal seems to be that our best psychological theory

¹⁰Another reading of their claim is that the natural language expression 'belief' ought to have a different meaning than the one it in fact has. If this is their suggestion, the proposal is quite odd given the use of English-speaker judgments as evidence. Whatever the meaning of the English expression 'belief' might be, surely the judgments of English-speakers regarding uses of the term is the primary source of data for giving a semantics for the expression. So if the goal is to give a semantic analysis of that expression, indicating that English-speakers are just plain wrong about the meaning of 'belief' makes the use of case-based judgments as evidence entirely unconvincing.

would be better served by making use of a term like ‘belief_s’ or ‘mind_s’ in explaining human cognition.

If this is the proposal on offer, two problems undercut the metaphysical claim they want to make about the constituents of belief. First, this suggestion presupposes that the language used to express theories in psychology has an externalist semantics. Second, even assuming that the scientific language of psychology has an externalist semantics, the use of speaker intuitions about cases like INGA* and OTTO’ as evidence for including terms that permit the admission of organism-external objects as constituents of the mind, is fairly strange. After all, English-speakers are speakers *of English*, not the technical language used for theorizing in psychology. But more importantly, if the suggestion is that a research program engaged in studying human psychology which includes extended-mind-friendly terminology has more explanatory power than one that does not, the arguments for such a view should appeal to the typical standards for evaluating scientific theories. The extended-mind-friendly theory must bear the hallmarks of good naturalistic inquiry: offering explanatory generalizations, making testable novel predictions, and integrating with other naturalistic domains, to name a few. Surely, that the extended mind hypothesis can explain the behavior of a single fictional individual more succinctly than more standard theories is (at best) weak evidence for the theory. Especially if it does so at the expense of much more mundane (real-world) cases (Adams & Aizawa, 2010).

This sort of argumentation trades on an assumption about the nature of linguistic meanings, namely that the meanings of the expressions we use determine their truth-conditions. This assumption carries with it an implicit demand for metaphysical consistency. Our judgments about the truth-conditions of the expressions we use, according to this demand, ought to remain ontologically consistent across contexts. This insistence permits leveraging cases like INGA*, that ostensibly indicate the paradigmatic truth-conditions of a particular expression (as exhibited by speaker judgments about the term of interest) to make metaphysical proclamations *via* supposedly structurally similar cases like OTTO’. The general thought behind this argumentative strategy is that there is some core, truth-conditional meaning to the natural language expressions we use. These meanings, while sometimes reflected in the judgments of competent speakers, require precisification by way of imaginative counterexamples that enable us to push the limits of those truth-conditional meanings. Of course, I think this line of reasoning is predicated on a mistaken assumption of natural language meanings.

The metaphysician then owes us a response to the problems for the externalist program, or she is forced to admit that her investigation does not pertain to natural language expressions and their meanings (which many, like Clark and Chalmers seem happy to admit). But this admission presents a further problem for the use of Conceptual Analysis, given the use of language as a tool for investigation. If the language the metaphysician uses to describe her intuition-pumping cases is not a natural language, then she owes us a justification for thinking meanings in this language are indicative of what exists.

A promising response to this challenge cites naturalistic methodology, appealing to the practice of theory building in the sciences. Naturalistic inquiry seeks to understand the world. The products of such inquiry are the theories that describe the world, which make use of terms invented for that purpose. Because the languages used to state our best scientific theories are designed to perspicaciously describe the world, so the metaphysician might argue, the terms of those theories have an externalist semantics. If the language used to investigate ontological questions is a scientific language, then the epistemic credentials of naturalist methods of inquiry provide (at least some) justification for thinking scientific terms “cut nature at the joints.” According to this naturalistic appeal such languages, and the meanings of the expressions in them, are useful in settling disputes about what there is.

However, even if this naturalistic appeal is successful, and indeed the languages invented to express our best scientific theories are reliable indicators of what there is, this provides (at best) little support for the extended mind hypothesis. The relevant question for metaphysical inquiry, under the guise of this naturalistic appeal, is whether a theory with the metaphysically interesting term admits to more explanatory successes with regard to the data *in the domain of inquiry* than a theory with no such term. If naturalistic inquiry yields a theory of human cognition that contains a term ‘mind’ whose extension included entities beyond the boundaries of the human organism, then we’d have good cause to endorse the extended mind hypothesis. Unless that domain is the naturalistic study of language, the evidential import of cases meant to pump intuitions (as described in a natural language) is (at best) insignificant. What matters is whether the research program that implies the existence of the proposed entity is successful along the dimensions that matter for the purposes of science. Such success is hard fought, over decades of investigation with tested methodologies, not the product of imaginative stories that exploit the flexibility of natural languages and the

speakers that understand them.

In the next section I'll show that an analogous concern arises in the metaphysics of *causation*. The hypothesis that human *minds* (as distinct from the meaning of 'mind' or the content of MIND) extend beyond the skull is (errantly) motivated by the intuitions garnered from Conceptual Analysis, solicited by the use of vignette's like OTTO and INGA. As illustrated above, such inquiry is ill-suited for understanding the nature of minds, even if it might have some (limited) utility in understanding the meaning of the English expression 'mind'. Only by assuming the troubled externalist hypothesis (E) can this utility be applied to address the ontology of minds. Metaphysical debates pertaining to causation likewise leverage the intuitions of natural language speakers to render verdicts about metaphysical theories. The same troubles that loom for the extended mind hypothesis in assuming (E) apply to the metaphysics of causation. The point I intended to underscore in the next section by highlighting to analogous character of these worries is that analogous problems will arise in any domain that make use of Conceptual Analysis to adjudicate between various metaphysical proposals.

5 Causation

As a topic of metaphysical investigation, study regarding the nature of *causation* has as storied a history as any, dating back to Plato's dialogues. Much more recently, reductive accounts of causation favor a counterfactual analysis indebted to Lewis (1973a). The primary goal of this research program seems to be one of modification, adding to and adjusting elements of Lewis' core insight to deal with the errant predictions of the theory. In this section, I'll begin by detailing the counterfactual account offered by Lewis as a means of highlighting the centrality of the externalist hypothesis (E) to theorizing in the metaphysics of causation. In characterizing the truth-makers for causal claims, Lewis makes some (implausibly) bold posits about what exists. As I'll show, Conceptual Analysis plays a central role in the development of Lewis' theory, guiding metaphysical inquiry into the nature of causation. However, if (E) is troubled for the reasons rehearsed in §2, the justification for endorsing Lewis' view, and the various ontological commitments carried in tow, is likewise troubled.

Lewis' proposal is to explain causal facts, as the truth-makers of causal claims, in terms of counterfactual dependence, thereby reducing such facts to

modal facts about possible worlds. For one event to cause a later event, the latter must counterfactually depend on the former. Formally, the proposal is this:

- (C) For any two distinct events c and e , c causes e *iff* there is a set of events (d_1, d_2, \dots, d_n) such that if c had not occurred, then d_1 would not have occurred; and if d_1 had not occurred, then d_2 would not have occurred; \dots and had d_{n-1} not occurred, then d_n would not have occurred; and had d_n not occurred, then e would not have occurred.

This analysis is counterfactual insofar as the right-hand side of the biconditional is a (series of) counterfactual conditional(s). Actualized events c and e stand in a causal relation on the condition that if (counter to fact) c had not actually taken place, then (roughly) neither would e . (This is a slight misstatement of the analysis given above, which is an *ancestral* version of direct counterfactual dependence—the reason for this difference will be explained below.)

The success of this account in analyzing causation depends on the manner in which it treats counterfactual conditionals. The familiar data-point is that counterfactual conditionals contrast with more standard conditionals, in that their truth-conditions are not obviously systematic. Contrastingly, the standard material conditional has well understood truth-conditions, and is always true whenever the antecedent of the conditional is false. If the conditionals in \mathcal{C} are taken to be *material* conditionals, then any actual event c would be the cause of any other event at any point in history, because the conditional ‘if c had not occurred, then e would not have occurred’ would never be false for any event e (as a result of the false antecedent). The familiar point is that counterfactuals can be (but are not always) true when their antecedents are false. Thus a successful reduction of *causal* facts to counterfactual dependence requires a compelling analysis of *counterfacts* (so to speak).

The account of counterfactuals given by Lewis (1973a)—and Lewis (1973b)—understands counterfactuals via possible worlds and the relations that hold between them. To quote:

Given any two propositions A and C , we have their counterfactual $A \Box \rightarrow C$: the proposition that if A were true, then C would also be true. The operator $\Box \rightarrow$ is defined by a rule of truth, as follows.

$A \Box \rightarrow C$ is true (at world w) iff either (1) there are no possible A -worlds (in which case $A \Box \rightarrow C$ is *vacuous*), or (2) some A -world where C holds is closer (to w) than any A -world where C does not hold. (Lewis, 1973b, p. 559)

The thought behind this technical account of counterfactuals is based on the idea that, given a plenitude of possible worlds, any set of these worlds constitutes a proposition (Lewis, 1973a, p. 556, note 3). Thus, $A \Box \rightarrow C$ is a claim about how two sets of possible worlds are related, the A -worlds and the C -worlds. These are the worlds at which for any sentence whereby A (or C respectively) is the sentence's (propositional) meaning, that sentence is true at each world that is a member of A (or C). The claim $A \Box \rightarrow C$, made at some possible world w , is true at w just in case the A -world closest to w is also a C -world. Put another way, there are a bunch of worlds where A is true, a bunch of worlds at which C is true, and a bunch of world where both are true. Given that some worlds can be "closer" to others, if one of those worlds where both A and C are true is closer to the world we care about (the world w at which $A \Box \rightarrow C$ is being evaluated/uttered) than any world where A is true and C is false, the counterfactual $A \Box \rightarrow C$ is true (at the world we care about, w).

Given this analysis of conditionals like $A \Box \rightarrow C$, one wants to know what makes a possible world *closer* to another. Lewis suggests that this relation is one of similarity, whereby two worlds are closer to each other than some third world if they are more similar to each other than either is to that third world. For the purposes of Lewis (1973a), he leaves the notion of similarity (and closeness) undefined, offering only a suggestive analysis of what similarity must (not) be like (Lewis, 1973a, pp. 559–560). Naturally much has been made about the nature of the similarity relation and the manner in which one is to weigh the various features of possible worlds in determining their proximal properties (cf. Bennett, 1974; Fine, 1975; Lewis, 1979, 1986b). Assuming a plausible analysis of similarity is available, the truth-conditions given above for sentences like ' $A \Box \rightarrow C$ ' offer a compelling account of counterfactuals, and thereby, causation.

Before we proceed further into the role of Conceptual Analysis in addressing the merits of Lewis' counterfactual analysis of causation, one should note the manner in which metaphysical inquiry has been guided thus far. In order to underwrite the *truth* of a given counterfactual claim, Lewis posits a plenitude of worlds that can be ordered along dimensions of similarity. Like-

wise, for any causal claim, the truth of that claim is understood in terms of the possible worlds that support the (truth-conditionally) related counterfactual. This process of ferreting out what domain entities are required to accurately codify the truth-conditions of expressions is a prevailing metaphysical methodology deployed in the field, and in this light clearly guided by linguistic meaning. One might deny that this methodology pertains to language at all (as Lewis claims, though this seems untenable), but its worth noting here the tight relationship between truth, the bearers of linguistic meaning (e.g. propositions), and the method used in metaphysical inquiry.¹¹

Returning to Lewis' account of causation by way of example, consider the sentence, uttered in the actual world ($w_{@}$):

- (13) If McCain had chosen a different running-mate, then Obama would have lost the 2008 presidential election.

Ignoring syntactic differences, this sentence can be put in a form amenable to Lewisian analysis:

- (13') [McCain chose a different running-mate] $\Box \rightarrow$ [Obama loses the 2008 presidential election].

Let's suppose that the English expression in (13) is true *iff* the (psuedo-English) expression in (13') is true. (This, of course, is not obvious, given that these expression have different syntactic structures, and one contains terms that are not familiar to English-speakers.) On Lewis' analysis of (13'), the meaning of 'McCain chose a different running-mate' is some proposition, as a set of worlds—call that proposition M . Likewise, the meaning of 'Obama loses the 2008 presidential election' is some proposition—call it O .¹² As such, (13) is true just in case ' $M\Box\rightarrow O$ ' is true (at $w_{@}$). If we consider the worlds at which McCain chooses a different running mate (someone other than Palin), we want to find the closest-to- $w_{@}$ world where McCain also wins the election (i.e. where Obama losses the election). Call that world w_{mo} . Still considering just the worlds at which McCain chooses someone other than Palin, we now

¹¹It is difficult to overstate how prevalent the appeal to sentences, truth, grammar, translation, and other linguistic properties is in the field. Consider the following non-exhaustive list canonical works that engage in such inquiry under the guise of metaphysics: Plantinga (1974); Jackson (1982); Lewis (1986a); Kim (1998); Smart (1963); Wiggins (1980); Kripke (1980); Hawthorne (2002); Thomasson (1999); van Inwagen (1990).

¹²Or more cautiously, the meaning of this expressions determines a unique set of worlds.

want the closest-to- $w_{@}$ world where Obama nonetheless wins. Call that world w_{mq} . Given these two worlds, if w_{mo} is closer to $w_{@}$ than w_{mq} is (to $w_{@}$), then (13) is true (and otherwise it is false). The account gets the correct result just in case the truth (or falsity) of (13) accords with the proximity of $w_{@}$ to w_{mo} and w_{mq} .

Further this analysis underwrites (let's suppose) the causal fact that McCain's choice caused Obama to win the election. The counterfactual account of causation indicates that some event c causes e just in case, if c did not occur, then e would not have occurred. Assuming that 'McCain chose Palin as a running-mate' and 'Obama wins the 2008 presidential election' describe different events, we can take the counterfactual analysis of (13) and apply it to an analysis to the causal fact that McCain's choice of Palin as a running-mate caused Obama to win the 2008 presidential election. This causal fact, according to the Lewisian analysis, reduces to the modal fact that $M \Box \rightarrow O$. Supposing c is the event described by 'McCain chose Palin as a running-mate', this event does not occur in all and only the M -worlds indicated earlier—those worlds in which McCain chooses *someone else* as a running-mate.¹³ Likewise, assuming 'Obama wins the 2008 presidential election' describes event e , this event fails to occur in all and only the O -worlds, since those are the worlds in which Obama loses the election. The counterfactual 'if c had not occurred, then e would not have occurred' is just the counterfactual $M \Box \rightarrow O$. In this way, the causal fact that McCain's choice of running-mate caused Obama's victory is reduced to modal facts about the proximity of M -worlds and O -worlds to the actual world. Assuming McCain's choice of running-mate did cause Obama to win the election, the (simple) counterfactual analysis of causation yields the correct prediction.

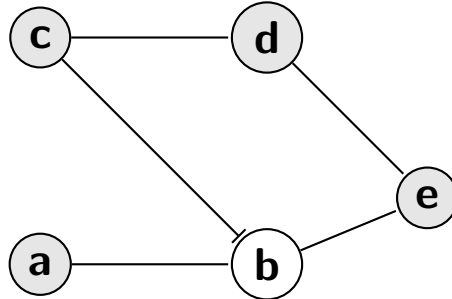
But as we noted above, the counterfactual analysis of causation offered by Lewis is slightly more complicated than the one we've rehearsed thus far. The discussion above reduces the casual relation between two events into the counterfactual dependence of *one* event on another. The (causing) event c causes the (effect) event e just in case c 's non-occurrence entails e 's non-occurrence. But this leads to (purportedly) counter-intuitive results.

Consider the diagram below, which represents the activation pattern of a neural network. Each circular node in the diagram represents a neuron, connected *via* the linear activation vectors passing from left to right in temporal

¹³This ignores, for ease of explanation, the possibility that McCain does not choose a running-mate at all.

order. The gray nodes are those that have activated, and the white nodes are those that remain unactivated. The normal arrows indicate the transmission of an activation signal stemming from an activated neuron, while the flat-headed arrow indicates a canceling signal that deactivates a node in the presence of a (distinct) activation signal. Reading the diagram, there are two causal paths, both of which bring about the activation of e . One originates from a and the other originates from c . When c activates it sends two signals: one runs down the connection with d , and the other (deactivation) signal runs down the connection to b . That deactivation signal cancels the signal originating at a , indicated by the whiteness of node b .

Figure 1: Neural Diagram



The contention is that this kind of case serves as a counterexample to the simple counterfactual account of causation. On the simple account, the fact that c 's firing causes e 's firing is reduced to the modal facts about the proximity of possible worlds in which c does not fire, to the actual world (where events unfold as indicated in figure 1). The closest such world, the one *most similar* to the actual world, is the one where c does not fire, yet a does. The reasoning here is that such a world is closer to the actual world than any other, since it instantiates the greatest number of events that hold in the actual world. Of course, at such an a -firing-world (where c fails to fire), e fires as well. As such, the simple counterfactual, that if c does not fire, then e does not fire turns out to be false. Likewise, the causal claim that c 's firing causes e 's firing turns out to be false. Ostensibly, this is counterintuitive, since there is a clear causal connection between c and e in the actual world.

Lewis' *ancestral* counterfactual proposal solves this problem. Since the ancestral notion (as indicated in (C)) requires us to evaluate *two* counterfactuals, given the intervening event of d 's firing, this permits us to isolate the conditions that make the counterexample troubling, and yield the correct

verdict about the causal claim that c causes e . The counterfactual $C \square \rightarrow D$ turns out true, since the closest non- c -world is also a non- d -world. Holding fixed all other events, when c fails to fire, d also fails to fire. The counterfactual $D \square \rightarrow E$ also turns out true on this proposal. Because we only need to consider the events moving forward from the time of d 's firing, when we consider the worlds at which d fails to fire, we are not required to retrodict the non-firing of c . Since we hold fixed all past events, including the firing of c which would prevent the firing of b , we maintain that b does not fire. Thus, the closest non- d -world is also a non- b -world, and hence a non- e -world. Given this analysis, both counterfactuals ($C \square \rightarrow D$ and $D \square \rightarrow E$) turn out to be true, and thus the causal fact that c 's firing causes e 's firing is predicted by the account.

The ancestral counterfactual account cannot, however, accommodate cases of *preemption*. These are cases that are structurally similar to the case above, but in such cases the intervening causal events are absent. Here we return to the relevance of ROCKS that appears in §3.1:

[ROCKS]

Billy and Suzy throw rocks at bottles. Suzy throws first, or maybe throws harder. Her rock arrives first. The bottle shatters. When Billy's rock gets to where the bottle used to be, there is nothing but flying shards of glass. (Lewis, 2000, p. 184)

Much like the neural case above, there are two distinct causal chains, both of which lead to the same terminating event. In ROCKS however, there are no intervening events on which the ancestral analysis can pivot, yielding the (ostensibly) counterintuitive result that Suzy's throw did not cause the bottle to shatter. If Suzy fails to throw her rock, Billy's rock shatters the bottle. Thus, the closest world at which Suzy fails to throw is a world at which the bottle nonetheless breaks.

The modern history of debates about causation is aptly described as offering criticisms of, and responses to, case-based counterexamples to Lewis' initial proposal, in much the way ROCKS purportedly does.¹⁴ Thus, it would be helpful at this juncture to ask what Lewis' analysis is supposed to analyze, and whether case-based intuitions aid in such an analysis. Lewis' account is clearly meant to offer up (\mathcal{C}) as an explanation of *something*, but what are those *explananda*? As we saw with the discussion of extend minds, there are

¹⁴See Paul & Hall (2013) for discussion.

a few possible candidates: the (use of the) English expression ‘cause’, the contents of a concept CAUSE that humans use to think about things/events as such, the content of an idealized term used for the purposes of science ‘cause*’, or the nature of the mind-external relation *cause* that binds *events*.

Consider the first candidate for the *explananda* of this account, the semantics for the English expression ‘cause’, as exhibited by the judgments of competent English-speakers. After all, cases like ROCKS are meant to solicit intuitions about causal sentences from the readers of those cases. Perhaps the purpose of this analysis, and the cases that inform it, is to give a semantics for the English expression ‘cause’. And given the (adventurous) assumption that the meaning of ‘cause’ is exhausted by the content of the human concept CAUSE, the first two options can be collapsed into one. Lewis addresses this question in a footnote, indicating that his proposal regards causal *facts*, not linguistic objects Lewis (1973a, p. 556, fn. 3). Further, Collins et. al. echo this goal in their introductory contribution to a prominent volume on the metaphysics of causation. In fact, they indicate that the central misstep of a competing analysis defended by Davidson (1967) is the focus on sentences instead of propositions (Collins et al., 2004, p. 17). They insist that the evidence brought to bear by dissecting cases, in the manner above, informs us about *propositions* and causal *facts*, not merely linguistic expressions that invoke the term ‘cause’.

Such a response lumps together the remaining two proposals. After all, if the world has a causal structure that we are privy to, and cases like ROCKS can serve as catalysts in uncovering that structure, the resulting term ‘cause*’ defined by way of (\mathcal{C}) ought to refer to those causal relations. However, insisting that human intuitions about cases like ROCKS inform us about the nature of propositions, and some underlying relation *cause* that constitutes (or is reduced to modal facts about) those propositions, presupposes that the English expressions used in describing these cases enable English speakers to grasp the propositions Lewis’ theory is meant to explain. Purportedly this is accomplished insofar as what an English speaker understands in comprehending the case are the truth conditions of the expressions used therein, and thereby determining what propositions are expressed. This assumption commits Lewis to the troubled externalist hypothesis (E). More importantly, even if we accept this (troubled) externalist assumption, to deny that this methodology is engaged in semantics is odd given the work propositions are meant to do. After all, if we take the relation *cause*, as it manifests in propositions, as the *explananda* of our theory, given that one of the primary philo-

sophical jobs of propositions is to serve as the meanings of natural language expressions, analyzing the purported modal facts about these propositions *via* cases must thereby be a semantic project.

If propositions (or their constituents) are the object of explanation, Lewis' theory gives us the truth conditions for the expressions that convey such propositions in terms of a proposed proximity relation that holds between possible worlds. One can *claim*, of course, that such a theory gives the truth-conditions of certain classes of expressions and thereby indicating what propositions they convey, yielding metaphysical verdicts about the nature of *causation*. But given that propositions, on the externalist proposal, are (inextricably related to) the meanings of natural language expressions, and that the cases used to ostensibly inform us about these propositions are presented using a natural language, the meanings of expressions like 'cause' play a central role in providing such information.

Cases like ROCKS are proposed to undermine the Lewisian hypothesis about the (reductive) nature of *causation*, because the view yields the wrong truth-value for the proposition (purportedly) expressed by 'Suzy caused the bottle to break', with respect to the world described in the case. Ostensibly, this is because the proposition that Suzy did not throw a rock, and the proposition that the bottle did not break, do not bear the right sort of relation to the world described in the case. But, we only possibly come to know this by having an intuition of a certain sort, because we can grasp which propositions are relevant for the purpose of such inquiry given the case-description—and *that* (according to the externalist) is a result of our ability to comprehend the English language expressions 'Suzy caused the bottle to break', 'Suzy did not throw a rock', and 'the bottle did not break'. Thus, it must be, that in giving the truth-conditions for the proposition that Suzy caused the bottle to break, Lewis' theory also gives us the truth-conditions for the English expression 'Suzy caused the bottle to break' whose meaning determines the proposition the theory is meant to analyze. Otherwise, our intuitions regarding ROCKS would have no bearing on the viability of Lewis' proposal. Thus, Lewis must be offering a semantic proposal about the meaning of the English expression 'cause', under the assumption that a semantics for this expression is given by, and indicative of, worldly objects and their (modal) properties.

Notice again the methodology made use of here, in giving an account of *causation*. On the assumption that externalism holds for natural languages, investigating the meanings of causal expressions like 'Suzy's throw caused the bottle to break' has led us to a number of metaphysical conclusions. Lewis'

contention is that his analysis of counterfactuals is the most parsimonious, and that this licenses some robust metaphysical conclusions (Lewis, 1986a, 1979). This analysis requires that we quantify over certain kinds of objects, like *events* and *possible worlds*. Since the truth-conditions for counterfactuals require a plenitude of possible worlds as their satisfiers, possible worlds must occupy the domain—a domain that, according to the metaphysician constitutes what there is. Unless a more metaphysically modest account of the meanings of counterfactual claims can explain the judgments of English language speakers, we can conclude that possible worlds exist. Further, that a proximity measure is required to give an adequate semantics for causal expressions, those possible worlds must stand in relations of similarity (again, in the absence of a better semantics for ‘cause’).

For the metaphysician, this is how metaphysical investigation proceeds, by consulting our capacity to understand expressions in a given language, and on the assumption that this understanding determines what the world must be like for expressions in that language to be true, proclaim what must exist if any (particular) sentences in that language are to be true. Many might balk at the idea that language is the focus of metaphysical methodology. But in the absence of rather bold assumptions about the relationship between the human linguistic competence and the conceptual system—and further assumptions about the nature of conceptual content—such reticence remains unmotivated.¹⁵ In the case we’ve been focusing on, for any causal expression to be true, there must be things in the domain like *events*, *possible worlds*, and a *proximity* relation that holds between triplets of worlds. But again, this method of investigating metaphysical questions is only viable if the language used for ontological investigation has an externalist semantics. To that end, natural languages seem ill-suited.

The metaphysician defending Lewis’ proposal could insist that the conception of *causation* under investigation is *not* the one that serves to underwrite the meanings of the natural language expressions containing ‘cause’, holding that some more refined, and ontologically distinguished conception is the subject of investigation. This conception is manifest in propositions; those things that are determined by the meanings of expressions in language. But, either this is a normative claim about what the natural language word

¹⁵The metaphysician that conceives of natural languages as mind-external objects might be able to skirt this worry, but such a conception of language renders the investigation of basic facts about the human linguistic capacity hopelessly fraught. See Vogel (2016) for focused arguments to this effect drawing from Chomsky (1986).

‘cause’ should mean, or it is a descriptive claim about what a precisified term in some ontologically privileged language means. That is, if the metaphysician denies that their project is “merely” giving a description of the meaning for the natural language expression that competent speakers of English use, then whatever truth-conditions are on offer as the meaning of the relevant term, those truth-conditions are not those of the English expression ‘cause’. This is to insist that there is some ontologically privileged language which contains a term, let’s call it ‘cause*’, whose meaning denotes the genuine notion of *causation* that these discussions of neural networks, and stone-throwings are meant to undercover.

But if metaphysicians are in the business of studying such a notion, the utility of pumping intuitions about cases like ROCKS is at best opaque. On the assumption that the (troubled) externalist thesis for natural languages holds, the purpose of consulting the case-based intuitions of English-speakers is somewhat clear, insofar as such individuals speak the language in which we ask them questions like “Is the sentence ‘Suzy’s throw caused the bottle to break’ true given ROCKS?” But if the term whose underlying propositional-constituent meaning is part of a non-natural, ontologically privileged language, then *English*-speakers judgments about sentences like ‘Suzy’s throw *caused** the bottle to break’ give us little ontological guidance. After all, such speakers are incapable of grasping the propositional meaning of a sentence in a language they do not understand,¹⁶ so their verdicts about cases will fail to involve the proposition up for analysis. Likewise, if the defender of this methodology wants to insist that there is some interesting scientific notion of *cause* that is useful for ontological investigation, such a philosopher must explain how stories told in a natural language like English make use of such a conception. Whether or not the languages used to express our best scientific theories need a term ‘cause*’ is a potentially interesting and philosophically fruitful hypothesis (cp. Frisch, 2014). But the plausibility of that proposal should live or die by the explanatory benefit wrought by the theories that include such a term. And surely, whether or not (say) physics needs a term ‘cause*’ cannot be decided by the linguistic judgments of English-speakers

¹⁶At least if the attempted means of grasping such a proposition is simply the comprehension of sentences expressed using a language they do not understand. Assuming there are propositional meanings, as an English speaker I can surely grasp the proposition that the cat is on the mat. But I cannot grasp that proposition by comprehending the French sentence ‘Le chat est sur le tapis’ since I am not a French speaker. *Mutadis mutandis* for the proposition expressed by using the term ‘cause*’.

regarding cases like ROCKS.

The metaphysician has two choices in adopting the methodology of Conceptual Analysis: either 1) she accepts that her project involves giving a semantics for natural language expressions, or 2) she insists that her investigation pertains to some ontologically privileged language that contains the term of interest. The first option rescues her use of intuition-pumping cases in ontological investigations, but commits her to the applicability of (E) to natural languages. As we have seen in §2 this is a risky commitment. The second option avoids the risky theoretical commitment (in exchange for a somewhat less risky theoretical commitment regarding the semantics for scientific languages), but still comes at a cost. Namely, she sacrifices the methodology that pumps intuitions of natural language speakers. For metaphysical investigations of *causation* the dilemma is profound, given that much of the literature makes use of natural language intuition-pumping cases as evidence for a particular theory about *causation*, while (at least implicitly) assuming without defense that (E) is true for natural languages.¹⁷

As was the case with ‘belief’ and the extended mind hypothesis, the argument for using case-based intuitions to supply evidence for a given theory about *causation* relies on the externalist assumption. In the wake of the evidence outlined above, the metaphysician investigating *causation* might insist that while the flexibility (or polysemy) of natural languages make them poor investigatory tools, the language she has in mind is rigid and precise in the way natural languages are not. One wants to know then why we should think that this precise language cleaves to the structure of the world. If this language is the product of research in a scientific domain, constructed for the purpose of expressing theories in that domain, the justification for using such a language in ontological investigation is parasitic on the epistemic credentials of naturalistic methodologies. But once she has accepted that the notion of interest is not one that serves as the meaning of a natural language expression, she undermines her argument for making use of speaker intuitions as evidence for a particular hypothesis about the notion in question. There might well be some terms *mind** and *belief** that, when added to the language of psychology, help to yield novel predictions and more useful explanatory generalizations. But there’s no reason to think English-speaker judgments

¹⁷For (a sampling of) examples of authors that make use of case intuitions as counterexamples: Hall (2000); Hitchcock (1996); Mackie (1974); Paul & Hall (2013) and citations therein; Schaffer (2005); Woodward (1984); Kim (1973); Menzies (1989).

about cases like OTTO' and INGA* are indicative of such success. Likewise, theories that use the term *cause** may yield some explanatory benefit across many scientific domains. But there is little reason to think that speaker judgments about cases like ROCKS are suggestive of such a benefit. Thus the metaphysician that makes use of Conceptual Analysis as a means of investigating metaphysical questions either owes us a response to the problems outlined for adopting an externalist semantics for natural language, or she must abandon her methodology in favor of a naturalistic one. Neither option bodes well for the current, prolific use of Conceptual Analysis to resolve metaphysical disputes.

6 Conclusion

There is an enduring tradition in philosophy of treating the meanings of natural language expressions as externalist. As with many traditions, the reasons for abiding this history often go unarticulated. As a consequence, the bold character of the externalist hypothesis is ignored, in favor of making equally bold proclamations about the nature of reality, and the nature of human minds. The flexibility of natural languages, combined with the naturalistic commitments that motivate these proclamations, countenance the difficulty in accepting these traditional assumptions about the nature of meaning. The metaphysician is thereby burdened to defend her methodology, or abandon it in favor of one more amenable to her naturalist commitments. However, if this newly adopted methodology invokes the use of scientific languages, insisting that the epistemic rigor of naturalistic investigation forges languages that “cut nature at the joints,” such a methodological shift precludes the use of Conceptual Analysis as a tool for ontological investigation, at least insofar as the case descriptions used therein are expressed in natural languages. Given the centrality of Conceptual Analysis to ontological investigation, excising this method requires a genuine revision to the practices of philosophers in addressing metaphysical questions, and a radical reexamination of the *explananda* in many philosophical domains. The upshot to this expulsion is the re-development of a field of research where the difference between the questions that have a hope of being addressed can be clearly marked from those that do not.

References

- Adams, F. R., & Aizawa, K. (2010). Defending the Bounds of Cognition. In R. Menary (Ed.) *The Extended Mind*. Mit Press.
- Bennett, J. (1974). Counterfactuals and Possible Worlds. *Canadian Journal of Philosophy* 4(2), 381–402.
- Black, M. (1952). The Identity of Indiscernibles. *Mind* 61(242), 153–164.
- Bromberger, S., & Halle, M. (1995). The Ontology of Phonology. *Studies in Logic and the Foundations of Mathematics* 134, 725–743.
- Burge, T. (1975). On Knowledge and Convention. *Philosophical Review* 84(2), 249–255.
- Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy* 4(1), 73–122.
- Chomsky, N. (1965). *Aspects of a Theory of Syntax*. MIT Press.
- Chomsky, N. (1977). *Essays on Form and Interpretation*. Elsevier.
- Chomsky, N. (1986). *Knowledge of Language*. Praeger Publishing.
- Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis* 58(1), 7–19.
- Collins, J., Hall, N., & Paul, L. A. (2004). Counterfactuals and Causation: History, Problems, and Prospects. In J. Collins, N. Hall, & L. Paul (Eds.) *Causation and Counterfactuals* (pp. 1–57). MIT Press.
- Davidson, D. (1967). Causal Relations. *Journal of Philosophy* 64(21), 691–703.
- DeRose, K. (1999). Contextualism: An Explanation and Defense. In J. Greco, & E. Sosa (Eds.) *The Blackwell Guide to Epistemology* (pp. 187–205). Blackwell.
- Eklund, M. (2006). Metaontology. *Philosophy Compass* 1(3), 317–334.
- Fine, K. (1975). Critical Notice of Lewis 1973. *Mind* 84, 451–58.

- Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review* 5, 5–15.
- Frisch, M. (2014). *Causal Reasoning in Physics*. Cambridge University Press.
- Glanzberg, M. (2011). Meaning, Concepts, and the Lexicon. *Croatian Journal of Philosophy* 11(1), 1–29.
- Hall, N. (2000). Causation and the Price of Transitivity. *Journal of Philosophy* 97(4), 198–222.
- Hawthorne, J. (2002). Deeply Contingent a Priori Knowledge. *Philosophy and Phenomenological Research* 65(2), 247–269.
- Hinzen, W. (2006). Internalism and Truth. *Mind & Society* 5, 139–166.
- Hinzen, W. (2007). *An Essay on Names and Truth*. OUP Oxford.
- Hitchcock, C. R. (1996). The Role of Contrast in Causal and Explanatory Claims. *Synthese* 107(3), 395–419.
- Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly* 32(April), 127–136.
- Kim, J. (1973). Causes and counterfactuals. *Journal of Philosophy* 70(17), 570–572.
- Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. MIT Press.
- Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.
- Lewis, D. (1973a). Causation. *Journal of Philosophy* 70(17), 556–567.
- Lewis, D. (1979). Counterfactual Dependence and Time’s Arrow. *Noûs* 13(4), 455–476.
- Lewis, D. (1986a). *On the Plurality of Worlds*. Blackwell Publishing.
- Lewis, D. (1986b). Postscript to ‘Counterfactual Dependence and Time’s Arrow’. In *Philosophical Papers* vol. 2. Oxford University Press.

- Lewis, D. (2000). Causation as Influence. *Journal of Philosophy* 97(4), 182–197.
- Lewis, D. K. (1973b). *Counterfactuals*. Blackwell Publishers.
- Mackie, J. L. (1974). *The Cement of the Universe*. Oxford, Clarendon Press.
- Menzies, P. (1989). Probabilistic causation and causal processes: A critique of Lewis. *Philosophy of Science* 56(4), 642–663.
- Pajitnov, A., & Pokhilko, V. (1984). *Tetris [Computer software]*. Lyon: Infogrames.
- Paul, L. A., & Hall, N. (2013). *Causation: A User's Guide*. Oxford.
- Peacocke, C. (1992). *A Study of Concepts*. MIT Press.
- Pietroski, P. M. (2005). Meaning Before Truth. In G. Preyer, & G. Peter (Eds.) *Contextualism in Philosophy* (pp. 255–302). Oxford University Press.
- Pietroski, P. M. (2008). Minimalist Meaning, Internalist Interpretation. *Biolinguistics* 4, 317–341.
- Pietroski, P. M. (2010). Concepts, Meaning and Truth. *Mind & Language* 24(3), 247–278.
- Pietroski, P. M. (2018). *Conjoining Meanings: Semantics Without Truth-Values*. Oxford University Press.
- Plantinga, A. (1974). *The Nature of Necessity*. Clarendon Press.
- Plantinga, A. (2003). *Essays in the Metaphysics of Modality*. Oxford University Press.
- Prior, A. N. (1967). *Past, Present and Future*. Clarendon Press.
- Rothschild, D., & Segal, G. (2009). Indexical Predicates. *Mind & Language* 24(4), 467–493.
- Schaffer, J. (2005). Contrastive Causation. *Philosophical Review* 114(3), 327–358.

- Shoemaker, S. (1984). *Personal Identity*. Blackwell.
- Sider, T. (2002). *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford: Clarendon Press.
- Sider, T. (2011). *Writing the Book of the World*. Oxford University Press.
- Smart, J. (1963). *Philosophy and Scientific Realism*. Routledge & Kegan Paul.
- Thomasson, A. L. (1999). *Fiction and Metaphysics*. Cambridge: Cambridge University Press.
- van Inwagen, P. (1990). *Material Beings*. Cornell University Press.
- Vogel, C. A. (2016). Lexical Flexibility, Natural Language and Ontology. *Croatian Journal of Philosophy* 16(1).
- Vogel, C. A. (under review). Realism and Lexical Flexibility.
- Wiggins, D. (1980). *Sameness and Substance*. Harvard University Press.
- Williams, B. (1979). Internal and External Reasons. In R. Harrison (Ed.) *Rational Action* (pp. 101–113). Cambridge University Press.
- Woodward, J. (1984). A Theory of Singular Causal Explanation. *Erkenntnis* 21(3), 231–262.