VOL. VII, NO. 41, NOVEMBER 2015

DISPUTATIO

INTERNATIONAL JOURNAL OF PHILOSOPHY

Petrus Hispanus Lectures 2014

Articles

Critical Notice

Book Reviews

Disputatio publishes first-rate articles and discussion notes on any aspects of analytical philosophy (broadly construed), written in English or Portuguese. Discussion notes need not be on a paper originally published in our journal. Articles of a purely exegetical or historical character will not be considered.

All submissions to *Disputatio* should be made to the managing editor by e-mail to disputatio@campus.ul.pt. Submitted manuscripts should be prepared for blind review, containing no identifying information, and be sent as a PDF or Word document attachment to the e-mail submission. The e-mail should have the subject 'Submission: [name of article]'. The body of the e-mail should include the author's name, institutional affiliation, address, and title of the submission. A short but informative abstract (approx. 150 words) at the beginning of the manuscript is also required, followed by 5 keywords. For more information on how to submit a manuscript, please read the instructions on our site. All authors will receive an e-mail confirmation of receipt of their submission.

All Submissions to *Disputatio* are triple-blind refereed: the names and institutional affiliations of authors are not revealed to the Editors, to the Editorial Board, or to the referees. Without the prior permission of the Editors, referees and Board members will not show to other people material supplied to them for evaluation. All published submissions have been anonymously reviewed by at least two referees.

Publishers may send book review copies to Célia Teixeira, *Disputatio*, Centro de Filosofia da Universidade de Lisboa, Faculdade de Letras, Alameda da Universidade, 1600-214 Lisboa.

Founded in 1996, *Disputatio* was published by the Portuguese Philosophy Society until 2002. From 2002, it is published by the Philosophy Centre of the University of Lisbon. *Disputatio* is a non-profit publishing venture. From 2013, *Disputatio* is published only online, as an open access journal.

PUBLISHED BY
VNIVERSITAS
Centro de Filosofia
OLISIPONENSIS

Editors: Teresa Marques and Célia Teixeira. Biannual publication. ICS registration number: 120449. NIPC: 154155470. Headquarters: Centro de Filosofia, Faculdade de Letras de Lisboa, Alameda da Universidade, 1600-214 Lisboa.

DISPUTATIO

INTERNATIONAL JOURNAL OF PHILOSOPHY

Vol. VII, No. 41, November 2015

Published by Centro de Filosofia da Universidade de Lisboa
ISSN: 0873 626X — Depósito legal n.º 106 333/96

Why Theories of Concepts Should Not Ignore the Problem of Acquisition

Susan Carey
Harvard University[1]

BIBLID [0873-626X (2015) 41; pp. 113-163]

Abstract
A theory of conceptual development must provide an account of the innate representational repertoire, must characterize how these initial representations differ from the adult state, and must provide an account of the processes that transform the initial into mature representations. In *The Origin of Concepts* (Carey 2009), I defend three theses: (1) the initial state includes rich conceptual representations, (2) nonetheless, there are radical discontinuities between early and later developing conceptual systems, (3) Quinean bootstrapping is one learning mechanism that underlies the creation of new representational resources, enabling such discontinuity. Here I argue that the theory of conceptual development developed in *The Origin of Concepts* constrains our theories of concepts themselves, and addresses two of Fodor's challenges to cognitive science; namely, to show how learning could possibly lead to an increase in expressive power and to defeat Mad Dog Nativism, the thesis that all concepts lexicalized as mono-morphemic words are innate. In response to Fodor, I show that, and how, new primitives in a language of thought can be learned, that there are easy routes and hard ones to doing so, and that characterizing the learning mechanisms in each illuminates how conceptual role partially determines conceptual content.

Keywords
Concept, concept acquisition.

[1] Reprinted courtesy of The MIT Press from *The Conceptual Mind*, edited by Eric Margolis and Stephen Laurence, Cambridge, MA: MIT Press, 2015 (https://mitpress.mit.edu/index.php?q=node/249683).

The *Petrus Hispanus Lectures 2014* were delivered by Professor Susan Carey at the University of Lisbon on May 27th and 29th 2014.

Disputatio, Vol. VII, No. 41, November 2015
Received: 05/11/2015

1 Introduction

The human conceptual repertoire is a unique phenomenon on earth, posing a formidable challenge to the disciplines of cognitive science. Alone among animals, humans can ponder the causes and cures of pancreatic cancer and global warming. How are we to account for the human capacity to create concepts such as CLIMATE, CANCER, ELECTRON, INFINITY, GALAXY and WISDOM? How do such concepts arise, both over history and in ontogenesis? Rightly, most attempts to provide such an account center on what makes concept attainment possible, but the literature on concept development adds a second question. Why is concept attainment (sometimes) so easy and what (sometimes) makes concept attainment so hard? Easy: some new concepts are formed upon first encountering a novel entity or hearing a new word in context (Carey 1978). Hard: others emerge only upon years of exposure, often involving concentrated study under metaconceptual control, and are not achieved by many humans in spite of years of explicit tutoring in school (Carey 2009). Considering what underlies this difference illuminates both how concepts are attained and what concepts are.

A theory of conceptual development must have three components. First it must characterize the innate conceptual repertoire—the representations that are the input into subsequent learning processes. Second, it must describe how the initial stock of representations differs from the adult conceptual system. Third, it must characterize the mechanisms that achieve the transformation of the initial into the final state.

The two projects of constructing a theory of concept acquisition and constructing a theory of concepts fit within a single intellectual enterprise. Obviously, a theory of concept acquisition must be consistent with what concepts *are*. But the relation between the two projects goes both ways, a fact that has played almost no role in the psychological literature on concepts (see, for example, the excellent reviews in Smith and Medin 1981, and in Murphy 2002). With the exception of developmental psychologists, cognitive scientists working on concepts have mostly abandoned the problem of characterizing and accounting for the features that enter into their learning models, often coding them with dummy variables.

This was not always so. For example, in theorizing about concepts, the British Empiricists made accounting for acquisition a central concern. They, like many modern thinkers, assumed that all concept learning begins with a primitive sensory or perceptual vocabulary. That project is doomed by the simple fact that it is impossible to express most concepts in terms of perceptual features (e.g., CAUSE, GOOD, SEVEN, GOLD, DOG...). In response, some theorists posit a rich stock of innate conceptual primitives, assuming that the adult conceptual repertoire can be built from them by conceptual combination. That is, they assume that the computational primitives that structure the adult conceptual repertoire and the innate primitives over which hypothesis testing is carried out early in development are one and the same set (e.g., Levin and Pinker 1991; Miller 1977; Miller and Johnson-Laird 1976). A moment's reflection shows this assumption is also wrong. For example, the definition of GOLD within modern chemistry might be ELEMENT WITH ATOMIC NUMBER 79. Clearly the theoretical primitives ELEMENT and ATOM are not innate conceptual features, as they arise in modern chemistry and physics only in the 18th and 19th centuries, after many episodes of conceptual change. (Of course, it is an open question whether ELEMENT and ATOM are definable in terms of developmental primitives; there are no proposals for possible definitions in terms of innately available primitives). Or take the features that determine the prototype structure of animal concepts (e.g., BIRD: FLIES, LAYS EGGS, HAS WINGS, NESTS IN TREES, HAS A BEAK, SINGS,...). Participants in studies provide just these when asked to list the features of birds. Furthermore, overlap in these features with others at this grain predicts judged similarity of birds to other animals, and overlap in particular values of them (e.g., beak type), as well as other features such as color and size, predicts prototypicality within the category of birds. That is, this feature space definitely underlies adult prototypicality structure. Prototype learning models assume that learning a new concept involves constructing a summary representation of a category in terms of such features, and then using this summary representation to probabilistically determine category membership. But a moment's reflection shows these models just help themselves to features that are not, for the most part, innate primitives—many are no less abstract nor no less theory-laden than the concept BIRD itself.

In a recent book (Carey, 2009, *The Origin of Concepts*, hereafter, *TOOC*), I take on the dual projects of accounting for conceptual development and characterizing the nature of human concepts. Towards a theory of conceptual development, I defend three theses. With respect to the initial state, contrary to historically important thinkers such as the British empiricists, Quine, and Piaget, as well as many contemporary scientists, the innate stock of primitives is not limited to sensory, perceptual or sensory-motor representations. Rather, there are also innate conceptual representations, embedded in systems of core cognition, with contents such as AGENT, OBJECT, GOAL, CAUSE, and APPROXIMATELY 10. With respect to developmental change, contrary to continuity theorists such as Fodor (1975), Pinker (2007) and many others, there are major discontinuities over the course of conceptual development. By 'discontinuity' I mean qualitative changes in representational structure, in which the later emerging system of representation cannot be expressed in terms of the conceptual resources available at the earlier time. Conceptual development consists of episodes of qualitative change, resulting in systems of representation with more expressive power than, and sometime incommensurable with, those from which they are built. Increases in expressive power and incommensurabilities are two types of conceptual discontinuities. With respect to a learning mechanism that achieves conceptual discontinuity, I offer Quinian bootstrapping.

Toward a theory of concepts that meshes with the picture of conceptual development in *TOOC*, I support dual factor theory (e.g., Block 1986). The two factors are sometimes called 'wide' and 'narrow' content. The wide content of our mental representations is partly determined by causal connections between mental symbols, on the one hand, and the entities to which they refer. To the extent this is so, all current psychological theories of concepts are on the wrong track—concepts are not prototypes, exemplar representations, nor theories of the entities they represent. However, contrary to philosophical views that deny that meanings are determined in any way by what's in the head (e.g., Dretske 1981, Fodor 1998, Kripke 1972/1980, Putnam 1975), *TOOC* argues that some aspects of inferential role are content determining (narrow content). The challenge for psychologists is saying what aspects of mental representation of

entities we can think about partly determine the meaning of concepts of those entities, and which are simply what we believe about those entities (sometimes called the project of distinguishing concepts from conceptions, Rey 1983). Facts about conceptual development constrain a theory of narrow content.

While the goal of *TOOC* was to explicate and defend the above three theses about conceptual development and sketch how they mesh with a dual factor theory of concepts, I also addressed Fodor's (1975, 1980) two related challenges to cognitive science—first, to show how learning can possibly result in increased expressive power, and to defeat the conclusion that all concepts lexicalized as monomorphemic words are innate. The key to answering both of these challenges, as well as to understanding conceptual discontinuities in general, is to show that, and how, new conceptual *primitives* can be learned. Conceptual primitives are the building blocks of thought, the bottom level of decomposition into terms that articulate mental propositions and otherwise enter into inference. Conceived of this way, there is no logical requirement that conceptual primitives cannot be learned.

Rey (2014) denies that the project is successful in meeting Fodor's challenges, as do Fodor (2010) and Rips and colleagues (Rips et al. 2008, 2013). Although I ultimately disagree, I appreciate many of the points these critics make along the way. These debates bring into focus how the projects of understanding conceptual development and understanding the nature of concepts, learning, and the human mind are intertwined. In this paper I lay out these debates on the interrelated issues of conceptual discontinuity, increases in expressive power, and Quinian bootstrapping and begin to sketch how they bear on our understanding of the nature of concepts. I show how new primitives can be learned, and how this fact bears on these debates.

2 The dialectic according to Fodor, Rey and Rips et al.

A kind of logical constructivism is at the heart of Fodor's and Rey's (and at least implicitly) Rips et al.'s dialectic. These writers, like many others, take expressive power to be a function of innate primitives, and what can—in principle if not in fact—be built from them

using the resources of the logic available to the learner. Expressive power is a logical/semantic notion. So long as the characterization of learning mechanisms is exhausted by specifying the set of innate primitives and the logical resources through which one builds new representations from those primitives, clearly one cannot increase expressive power by learning (Fodor 1980).

My response to this picture of learning and conceptual development is to argue that learning mechanisms can create new primitives, new primitives that cannot be constructed from antecedently existent primitives by logical combination, and thus increase the expressive power of the conceptual system. In addition, my concern is with how new primitives actually come into being; if there are processes that yield new primitives, then the question is whether such processes actually underlie the emergence of any given representation.

Fodor's (1975) second challenge to cognitive science is to defeat his argument for Mad Dog Nativism, that is, to defeat the argument that virtually all of the over 500,000 concepts lexicalized by monomorphemic words in the Oxford English Dictionary are innate. Rey (2014) lays out Fodor's argument as follows:

> Premise 1: (Hypothesis Confirmation). All learning is hypothesis confirmation.
>
> Premise 2: (Logical Construction) One can learn new concepts only by creating and confirming hypotheses formulated in terms of logical constructions from antecedently available primitive concepts.
>
> Premise 3: (Atomism). The concepts underlying mono-morphemic words cannot be analyzed as logical constructions of other concepts, primitive or otherwise. (Actually, Fodor says 'most' mono-morphemic concepts cannot be so analyzed, but for simplicity I will assume 'all' rather than 'most').
>
> Conclusion: (Innateness). In order to acquire a new concept lexicalized as a mono-morphemic word, one would have to confirm hypotheses already containing the concept to be learned. Therefore, no such concept can be learned.

TOOC answers this challenge by giving reasons to deny premises 1 and 2. My basic strategy has been to provide several case studies of transitions between conceptual systems in which the later one expresses concepts that are not logical constructions from the earlier one (Carey 1985, 2009; Smith, Carey and Wiser 1985; Wiser and Carey 1983). Sometimes this is because of local incommensurability, as in case studies of thermal concepts, biological concepts and electromagnetic concepts in the history of science, or concepts of matter/weight/and density in intuitive physics in childhood and the concepts of life and death in childhood). Sometimes it is because of developments within mathematic representations that increase expressive power without necessarily involving local incommensurability (as in case studies of the origins of concepts of integers and rational number).[2] *TOOC* then goes on to analyze how Quinian bootstrapping plays a role in transitions of both types.

The central issue dividing my views from the critics I focus on here is discontinuity. These critics deny the very possibility of conceptual discontinuities, as well as offering a positive view of conceptual development in terms of Premises 1 and 2 of Fodor's argument which they claim shows how conceptual development is possible without discontinuity. Rips and his colleagues suggest that claims for discontinuities are incompatible with claims that concepts are learned (Rips and Hespos 2011; Rips, Asmuth and Bloomfield 2013). Again, the key is understanding that, and how, new conceptual primitives can be learned. These critics argue that my proposal for a learning mechanism that can underlie conceptual discontinuity, Quinian bootstrapping, fails, partly through failing to confront a psychologized version of Goodman's new riddle of induction (Rey 2014, Rips et al. 2008).

With respect to Rips' and his colleagues worries that concept learning and concept discontinuity are incompatible, let me clarify what the debate is *not* about. The existence of conceptual discontinuity cannot entail that it is impossible for an organism to acquire some

[2] The case study of the construction of the integers is the focus of Rey's, Rips et al.'s, and Fodor's critiques. I will discuss whether this episode of conceptual development truly involves a discontinuity, and an increase of expressive power, when I turn to it in Sections 8 and 9 below.

later representations, given its initial state, except through maturation or magical processes that don't involve learning (e.g., being hit on the head). What is actual is possible. The mechanisms (there are many) that underlie the acquisition of our representational repertoire, in general, and our conceptual repertoire in particular, if they are learning mechanisms, are computational processes. At stake are premises 1 and 2 of Fodor's argument, which all of these critics explicitly or implicitly endorse. I agree that most of conceptual development consists of hypothesis confirmation, where the hypotheses are articulated in terms of already available concepts. Discontinuities arise in episodes of conceptual development where this is not the right model.

With respect to the positive proposal, Mad Dog Nativism requires that virtually all the 500,000 concepts lexicalized in English, plus those that will come to be lexicalized in the future, are innate, existing in some way in the infant's mind. This isn't comforting as a positive proposal that obviates the need for concept learning. A priori, it is highly unlikely that QUARK and CARBURETOR and FAX are innate concepts, existing is some kind of hypothesis space available for hypothesis testing. Noting this unlikelihood, Rey (2014) distinguishes between manifest concepts (those currently available for hypothesis testing and inference) and what he calls 'possessed' concepts (those that exist in the mind in some way, but are not currently available for thought, or those that can be constructed, by logical combination from that initial set). Rey defines possessed concepts as those that have the *potentia*l to be manifest. Here I use 'potential' concepts instead of 'possessed' concepts to express this notion. Nobody would ever deny that an actual manifest concept had the potential to be the output of some developmental process, and in the light of characterizations of those developmental processes, we can and do explore the representational repertoire it can achieve. Exploring the possible outputs of the learning mechanisms we investigate is an important part of characterizing these mechanisms. Calling the potential output of concept learning mechanisms 'possessed concepts' implies something stronger, that they exist somehow in the mind prior to becoming manifest. Of course, Premises 1 and 2 specify one way we can think about this stronger notion 'possession:' the innate primitives, along with the combinatorial apparatus of logic and language

constitutes a space of alternative hypotheses about which concepts apply in particular contexts (e.g., to support the meaning of a word), and this space exhausts the potential concepts that are attainable. The writers I am criticizing here assume that potential concepts constitute a space of alternatives, laying in wait to become manifest, and that manifestation consists in *being* or *being logically constructed* from these innately possessed primitives. These assumptions follow from premises 1 and 2 of Fodor's argument, the premises I deny.

3 Initial response

My project concerns manifest concepts. To reiterate, manifest concepts are those currently available to for thought, inference, and guiding action. The developmental primitives I study are those we can find evidence for in the baby's or animal's behavior. They must be available to support inference and action in order to be diagnosed, i.e., they must be manifest (currently available for thought). In what follows I argue that concept manifestation is where the debates about expressive power, conceptual continuity/discontinuity, and induction *actually* play out.

For any representational system we posit, we are committed to there being answers to three questions. First, what is the format of the symbols in the system; second, what determines their referents; and third, what is their computational role in thought. A worked example in *TOOC* is the evolutionarily ancient system of number representations in which the mental symbols are quantities (rates of firing, or size of populations of neurons) that are linear or logarithmic functions of the cardinal values of sets, which in turn are input into numerical computations such as number comparison, addition, subtraction, multiplication, division, ratio calculations, probability calculations, and others (see Dehaene 1997, for a book-length treatment of this system of numerical representations). We can only explore such systems with psychological methods that diagnose manifest representations. The project of *TOOC* is understanding the representational resources available as the child or adult interacts with the world, how these arise and change over development. These representations are the ones available for hypothesis testing, as input into further learning, and to play a computational role in thought.

And it is successive manifest conceptual systems one must analyze to establish qualitative changes (i.e., conceptual discontinuities).

In what follows I flesh out these points, explicating how *TOOC* attempts to answer Fodor's challenges to cognitive science. The issues include a characterization of the nature of learning (Fodor's first premise), the unjustified acceptance of the logical construction model as the only model of concept learning (Fodor's second premise), the misleading analogy of the totality of concepts ultimately attainable as a hypothesis space, the characterization of how primitives arise (both in cases where this is easy and in cases where this is hard), and the characterization of constraints of induction (and constraints on learning more generally, in cases where learning does not involve induction).

Let me begin with the premises in Fodor's argument that I deny. I first comment on why these premises matter and I then show why they are wrong.

4 Premise 2. Logical construction

The premise that all concepts must either be innate or buildable by combination from innate primitives through innate logical combinatorial devices is widely adopted within cognitive science. For example, the dominant theoretical project within the field of lexical development in the 1970s was to attempt to discover the lexical primitives in terms of which lexical items are defined, and to study the intermediate hypotheses children entertain as they construct new concepts from those primitives (see Carey 1982, for a review and critique). That is, it was just assumed that definitional primitives are innate. There I called this view 'piece by piece construction'; Margolis and Laurence (2011) call it 'the building blocks model'. Here, I will call it 'the logical construction model', in honor of Premise 2. In contrast, I argue (Carey 1982, *TOOC*) that computational primitives need not be innate. They can be acquired through learning processes that do not consist of logical construction from innate primitives.

One central issue is atomism. If many of the primitives in adult thought (e.g., the concepts expressed by words like 'dog' or 'cancer'), cannot be defined in terms of innately manifest concepts, then

they either must be innate primitives or it must be possible to learn computational primitives through some mechanism that does not consist of building new concepts by logical combination of antecedently available ones, and is not exhausted by confirming a hypothesis stated in terms of the to be acquired concept. I accept Fodor's arguments that most lexical concepts are definitional primitives.

Notice that the possibility one can learn new primitives matters to the question of expressive power of the system. The expressive power of a system of representations is a function of its atomic terms and combinatorial apparatus. The logical connectives and operators (sentential operators, modals, quantifiers) are not the only primitives that matter to expressive power. If DOG cannot be logically constructed from primitives, then acquiring the concept DOG increases expressive power of the system (see Weiskopf 2008). That is, nonlogical primitives figure into semantic/logical expressive possibilities as well as do logical ones. This is one reason that the question of whether one *learns* the concept DOG is so central to the debate between Fodor and his critics.

5 Premise 1. All learning is hypothesis formulation and testing

To evaluate this proposition we must agree upon what hypothesis testing is and what learning is. Bayesian models specify the essence of hypothesis testing algorithms. Hypothesis testing requires a space of antecedently manifest concepts, each associated with prior probabilities, and each specifying likelihood functions from any possible evidence to the probability that it supports any given hypothesis. Hypothesis testing then involves choosing among the alternative hypotheses on the basis of evidence. Fodor (1975, 2008) claims that all learning mechanisms reduce to hypothesis testing, at least implicitly. I agree that any learning mechanism that revises representations as evidence accumulates (e.g., associative mechanisms that update strengths of association, supervised learning algorithms such as connectionist back propagation) do indeed do so. However, as Margolis and Laurence (2011) point out in a reply to Fodor's 2008 book (*LOT2*), a cursory examination of the variety of attested learning mechanisms in the animal kingdom shows that the generalization that *all* learning mechanisms reduce to hypothesis confirmation is

wildly off the mark. Rote learning (memorizing a phone number), one-trial associational learning (e.g., the Garcia effect, the creation of a food aversion as a result of becoming nauseous some fixed time after having eaten a novel food, Garcia et al. 1955), and many other types of learning do not involve choosing among multiple hypotheses, confirming one of them, in the light of accumulating evidence. And as we shall see, such mechanisms have roles to play in creating new conceptual primitives.

Of course, the claim that these are learning mechanisms depends upon what one takes learning to be. Learning mechanisms share a few essential properties that allow us to recognize clear examples when we encounter them. All learning results in representational changes in response to representational inputs, where those inputs can be seen (by the scientist) to provide evidence relevant to the representational change. That is, learning is a computational process, requiring representational inputs that can be conceptualized as providing relevant *information*. Sometimes, as in the case of explicit or implicit hypothesis testing, the organism itself evaluates the information in the input with respect to its evidential status (as in all forms of Bayesian learning mechanisms). But other times, the learning mechanism is a domain specific adaptation that responds to information by simply effecting a representational change of relevance to the organism—an example being the learning mechanism that underlies the Garcia effect mentioned above. No further evidence is evaluated, so there is no hypothesis confirmation.

6 The relatively easy route to new representational primitives: domain specific learning mechanisms

The problem of acquisition arises in the case of any representation, conceptual or otherwise, that end up in the manifest repertoire of an animal. The literatures of psychology and ethology have described hundreds of domain-specific learning mechanisms that simply compute new representations from input, having arisen in the course of natural selection to do just that. Most of these representations are not conceptual ones, but considering how they are acquired shows that the learning mechanisms involved do not always involve hypoth-

esis testing, thus providing counterexamples to Premise 1. They also do not implement logical construction from primitives, and thus provide counterexamples to Premise 2. Considering how they work illuminates why it's a mistake to consider potential representations as a space of existent representations, ready to be *chosen among* or *built from* in a process of manifestation.

TOOC's example of an evolved domain-specific learning mechanism is that which underlies Indigo buntings' *learning* which part of the night sky indicates north. This matters crucially to Indigo buntings, for they migrate over 3500 miles each spring (north) and fall (south), and they navigate by the stars. Because the earth tilts back and forth on its axis, what part of the night sky indicates north changes radically on a 30,000 year cycle. Sometime not too far in the future, the north star will be Vega, not Polaris. Thus, it is unlikely that an innate representation of Polaris as the north star was created by natural selection, and indeed, Steven Emlen (1975) discovered the learning mechanism through which Indigo buntings create the representation of north that will play such a crucial role in their migratory life. The learning device that achieves this analyzes the center of rotation of the night sky, and stores the configuration of stars that can allow the bird to recognize the position of north from a static sighting (as it has to do every time it starts to fly during its migrations in the spring and the fall, and as it monitors its course).

This mechanism computes what it is designed to compute—nothing more, nothing less. It creates an essential representation in the computational machinery of Indigo buntings, the specification of north in the night sky. Of course, there is a prepared computational role for this representation, but the representation of north as specified by the stars must still be learned, and is an essential primitive in the computational machinery underlying Bunting navigation. Domain specific learning mechanisms of this sort are often supported by dedicated neural machinery that atrophies after its work is done, leading to critical periods. This is such a case; if a bird is prevented from seeing the night sky as a nestling, no amount of exposure to the rotating night sky later in life allows the bird to identify north, and the bird perishes.

This example is worth dwelling upon with respect to whether representations that can be achieved should be thought of as part of

an existing space of hypotheses, and whether the acquisition mechanism involves hypothesis confirmation or logical combination. Until the learning episode is completed, there is no manifest representation that specifies north in the night sky in the bird's mind. However, this learning mechanism can learn any of a very large number of star configurations constellations that could indicate north. Indeed, part of the evidence that this *is* the learning mechanism through which indigo buntings establish Polaris as the north star are planetarium experiments in which the night sky is made to rotate around an arbitrarily chosen part of the night sky while the birds are nestlings. The birds then use the north star so specified to set their course when it's time to migrate. Thus, there are a plethora of potential north stars. And clearly, one can investigate limits on the system (e.g., if stars were equally distributed throughout the sky, or if they were too densely packed to be resolved, or if the patterns of stars showed large scale repetitions, this couldn't work.) It is only with an actual representational/computational characterization of this learning mechanism that the space of potential north stars the Bunting could aquire representations of can be explored. Such is always the case.

What about hypothesis testing? I take the essential features of hypothesis testing to be two: (1) the learning mechanism must entertain alternatives, and (2) choice among them must be based on evidence. The space of potential representations of north that can be achieved by Buntings is in *no* way a hypothesis space. In no way does an Indigo Bunting's acquiring a representation of north consist of choosing among possibilities. Calling the possible specifications of north a 'hypothesis space' is wildly misleading. There is no initial set of possibilities, with associated priors, with likelihood functions associated with them. The animal never considers any possibility other than the output of the learning mechanism, and the animal has no way of testing whether the specification of north that is the output of the learning mechanism is actually NORTH. The bird simply computes it, and lives or dies by it.

This case is also worth dwelling upon with respect to the other issues on the table. Not only does this case not involve hypothesis formulation and testing, it also does not involve building a new representation out of primitives by logical combination. And since there is no induction involved, the issues of constraints on induction do not

arise. Of course, all learning mechanisms must be highly constrained to be effective, and characterizing real learning mechanisms allows us to understand the constraints under which they operate. This is a highly constrained learning mechanism; it considers only one kind of information to create a representation that has only one computational role. It is of no use to the bird in helping the bird learn what to eat, who to mate with, or where its nest is in a local environment.

Navigation is not a special case. There have been hundreds of such domain specific learning mechanisms detailed in the literatures of ethology and psychology, including the imprinting mechanisms that allow infants (animals and humans) to identify conspecifics in general and their caretakers in particular, mechanisms that allow animals to learn what food to eat (the Garcia effect just one of dozens of domain specific learning mechanisms through which omnivores like rats and humans achieve this feat), bird song learning, and so on (see Gallistel et al. 1991, for a review of four such domain-specific information-expectant learning mechanisms, and Gallistel 1990 for a nuanced discussion of the nature of learning).

In sum, the animal literature provides many examples of learning mechanisms designed to form new computational primitives, learning mechanisms that implicate neither logical construction from existing primitives (Premise 2), nor hypothesis testing and confirmation (Premise 1). One can (and one does) explore the space of possible outputs of these mechanisms, for this is one way they can be fully characterized and their existence empirically tested, but in no way is there a space of representations laying in wait, existing ready to be manifested, existing ready to be chosen among.

7 The relatively easy route to new conceptual primitives

The learning mechanism described above acquires a new primitive representation, a representation that allows the animal to identify north in the night sky, to guide navigation. One might argue it is not a new *conceptual* representation. Its format is surely iconic, and its computational role is both highly domain specific and sensori-motor. There are, however, learning mechanisms that similarly respond to inputs of certain types by simply creating new conceptual primitives, primitives that enter into representations with propositional format

and participate in the full productivity of language and causal inference. These domain specific concept learning mechanisms need not involve hypothesis testing, and do not involve constructing new concepts by logical combination. Take the Block (1986)/Macnamara (1986)/Margolis (1998) object-kind learning mechanism for example.[3] This learning mechanism is triggered by encountering a novel object (as specified by core cognition of objects) with obviously non-arbitrary structure. As Prasada et al. (2002) showed, there are several cues to non-arbitrary structure: the object has complex yet regular shape (e.g., symmetries, repetition), or there are multiple objects that share a complex irregular shape, or the object has functionally relevant parts, or the object recognizably falls under an already represented superordinate kind (e.g., kind of agent, kind of animal, kind of artifact). Core cognition contains perceptual input analyzers that are sensitive to cues to each of these properties of individual objects. Encountering an individual with one or more of these properties triggers establishing a new representational primitive that can be glossed SAME BASIC LEVEL KIND AS THAT OBJECT. Reference to the kind is ensured by representation of the surface properties of the individual or individuals that occasioned the new concept (and these represented surface properties get enriched and even overturned as bases of reference and categorization as more is learned about the kind). The content of the new concept depends upon the referent, the conceptual role provided by the basic level kind schema (psychological essentialism), and the conceptual roles provided by any superordinate kind schemas that the individual is taken to fall under (e.g., AGENT, ANIMAL, ARTIFACT, these in turn being constrained by their roles in different systems of core cognition or constructed theories).

Consider encountering a kangaroo for the first time. Such an encounter might lead to the formation of a concept KANGAROO that

[3] These writers discuss this mechanism as a *natural kind* learning mechanism (e.g., kinds of animals or kinds of plants), but I believe the domain of this mechanism is object kind representations (as opposed to object properties, individual objects, or the events in which objects participate). Roughly, kind representations are inductively deep, and kinds are construed in accordance with the constraints that constitute psychological essentialism in Strevens' (2000) sense. Artifact kinds fall under the domain of this mechanism as well as do natural kinds (Kelemen and Carey 2007).

represents animals that are the same basic level kind as the newly encountered one. No enumerative induction is needed; the concept is what Strevens (2009) calls 'introjected' into one's set of primitives. This concept, falling under psychological essentialism (as it is a kind concept), reflects the many constraints on kind concepts. That is, the conceptual role SAME KIND AS includes assumptions that something causes the non-random structure that triggered the formation of the new concept, that these underlying causes are shared by all members of the kind (now, in the past, in the future), that the surface properties that specify the individual that occasioned the new concept may not hold for all members, possibly not even typical members. Furthermore, the current guesses about the nature of the relevant causal mechanisms relevant to the creation of members of this kind, to determining their properties, and to tracing numerical identity though time, are taken to be open to revision. That is, there is no definition that determines membership in the kind; learners treat everything they represent about the kind up for revision (including, even that there IS a new kind—the individual we encountered might have been a mutant raccoon).

This mechanism creates new primitives, not definable in terms of other manifest concepts, and thus increases the expressive power of the conceptual system. The concept KANGAROO is not definable in terms of antecedently available primitives using the combinatorial machinery of logic. Before creating this concept, one could not think thoughts about kangaroos, just as before analyzing the center of rotation of the night sky and storing a representation of NORTH so specified, an Indigo bunting could not set or guide a course of flight toward or away from north. Of course the kind learning mechanism ensures that creating new primitives for kinds is easy; one need only encounter an individual that one takes to be an individual of a new kind, and store a representation of what that individual looks like. But this process involves neither induction nor hypothesis testing among a huge space of antecedently available innate primitives. The concept KANGAROO was not laying in wait in a system of representations available for selection by a Bayesian hypothesis testing mechanism, nor is it constructible by logical combination from antecedently available primitives.

Rey (2014) discusses the Margolis kind learning module, claiming

that it falls prey to Goodman's grue problem, just as Quinian bootstrapping does (see below). There are two answers to Rey's questions regarding constraints on induction in the Margolis kind learning module. First, as detailed above, there need be no induction. But, Rey asks, why are not kinds such as objects, animals, agents, Eastern grey kangaroos, kangadiles (kangaroos until year 2040, thereafter crocodiles), undetached kangaroo parts, or an infinitude of other kinds, possible glosses of SAME KIND AS THAT OBJECT, rather than the kind kangaroo? Why does the learner not form a concept of a particular individual (Oscar) instead of a kind?

Answering this question simply *is* an important part of the project understanding conceptual development. In the case of dedicated concept learning devices such as the object-kind learning device, the empirical project is specifying the constraints under which the system operates. That there is a dedicated kind concept acquisition device is an empirical discovery, and, like all learning mechanisms this one embodies strong constraints. It is a discovery that there is basic level in kind concepts, and it is a discovery that basic level kinds are privileged in kind concept learning (e.g., Rosch et al. 1976). It is a discovery that kind representations embody constraints derived from causal/functional analyses (see the work on psychological essentialism and the psychology of a causal/explanatory core to kind concepts: e.g., Gelman 2003, Keil 1989, Ahn and Kim 2000, Strevens 2000). And the existence and structure of systems of core cognition (in which the concepts AGENT and OBJECT are embedded), as well as innately supported systems of causal and functional analysis, are empirical discoveries, as is the fact that these constrain kind representations from early infancy (Carey 2009). These constraints do not rule out *ever* entertaining concepts for attended individuals. After all, some concepts that are not basic level are themselves innately manifest (e.g., AGENT) and are drawn upon as important parts of the constraints on the kind module. That is, AGENT is the content of a superordinate kind that constrains a newly formed basic level kind concept that falls under it. Others, such as subordinate and superordinate kinds, as well as stage and phase sortals like PUPPY or PASSENGER, are routinely manifested after basic level kind representations are formed (e.g., Hall and Waxman 1993). Still others are obviously entertainable (after all, Goodman and Quine did so, and we

all can join in). But these concepts simply are not the output of the dedicated basic level kind learning device discussed above. Furthermore, the child *can* also form a concept of a particular individual, even a newly encountered kangaroo. There is a dedicated learning mechanism for concepts of individuals, as well as for basic level kinds (but that is another story, one that has also been told; e.g., Belanger and Hall 2006). *Once* cognitive science has discovered the constraints under which actual learning devices operate, one can explore their possible outputs. The constraints posited are empirical proposals, falsifiable by demonstrations that they are easily violated. The empirical work strongly supports the existence of the basic level object kind learning module.

The basic level kind learning module creates new primitive concepts. Before a person has formed the concepts KANGAROO or SHOVEL, or concepts of any of infinitely many new kinds, he or she cannot think thoughts about the entities that fall under those concepts. This learning mechanism thus results in an increase in expressive power. However, like the cases of the dedicated learning mechanisms discussed in the ethology literature (those that yield representations of conspecifics, caretakers, the north star), there is an innately specified conceptual role for kind concepts, in this case given by the abstract concept KIND OF OBJECT and by the schemas of superordinate kinds embedded in core cognition and constructed theories that the learner assigns the new concepts to. Such already existing schema and conceptual roles are always part of the relatively easy route to new primitives.

8 The dual factor theory of representations with innate conceptual role

Dual factor theory applies straightforwardly to concepts in core cognition (AGENT, OBJECT…), indeed any concept with innate conceptual role and innate perceptual input analyzers that support identification of entities that fall under it. The innate perceptual input analyzers explain how symbols are causally connected to the entities they represent, and the innate conceptual role specifies the narrow content of the concept. In core cognition, and cases like the indigo bunting

representations of the azimuth, the innate conceptual role is never overturned—the narrow content of the representation of the north star that makes it a representation of north simply *is* the suite of sensori-motor computations supporting navigation it enters into.

The story for the Block/Macnamara/Margolis kind module is a little less straightforward. In concepts created by the kind learning device there are innate input analyzers that trigger the establishing of a kind representation (that identify objects with non-accidental structure) and that support the identification of superordinate schema provided by core cognition (KIND OF OBJECT, KIND OF AGENT...). These innate input analyzers are part of what provides the wide content of such concepts, as they trigger forming a representation of an entity in the world that is part of the wide content of the newly formed concept, as well as providing part of the causal connection between this wide content and the newly formed mental symbol. But there is no innate, un-overturnable, prepared conceptual role at the level of specific kinds. Even the initial superordinate schema the kind is subsumed under is revisable. However there is innate conceptual role for object kinds in general (i.e., given by psychological essentialism), and this specifies what sort of concept is in play and constrains its formal properties. This abstract conceptual role specifies part of the narrow content for kind concepts. As Block (1986) says, it determines the nature of the connection between symbols and the world, after a symbol is taken to be a symbol for an object kind.

9 The relatively hard route to new conceptual primitives

Quinian Bootstrapping is a learning mechanism that also creates new primitives, thus increasing the expressive power of the conceptual system. It differs from those learning mechanisms described above in that it did not arise through natural selection to acquire representations of a particular sort. Rather, it is one of the learning mechanisms that underlie the creation of representational resources that are discontinuous with (in the sense of being qualitatively different from, being locally incommensurable with, the representations of the same domain that were their input). It creates new conceptual roles, rather than merely creating new primitives for which there were prepared conceptual roles (as in the case in the easy route to

new primitives, see above). But once created, these new conceptual roles provide constraints on the concepts that will be learned, just as in the relatively easy route to new conceptual primitives.

TOOC takes a particular episode along the way to creating a representation of integers as a central worked example of conceptual discontinuity and of Quinian bootstrapping. I argue that this case involves an increase in expressive power, in that before the bootstrapping episode the child has no manifest concepts for natural numbers, and the process of construction of the first representations of new primitive concepts, those of a subset of the natural numbers, is not exhausted by defining them in terms of primitives antecedently available. Again, let me be clear. The increase in expressive power at stake here is an increase in the expressive power of manifest concepts available to the child. Obviously the total computational machinery available to the child has the capacity for this construction (what is actual is possible); just as the computational machinery of the child has the capacity to create representations of kangaroos in the easy route to new primitives.

Expressive power is a semantic/logical issue. Examples of questions about expressive power relative to number representations include whether arithmetic can be expressed in the machinery of sentential logic (provably no) and whether arithmetic can be expressed in the machinery of quantificational logic plus the principle that 1-1 correspondence guarantees cardinal equivalence (provably yes, if you accept Frege's proof). But the exploration of expressive power with such proofs is relevant to the question of how arithmetic arises in development *only* against empirically supported proposals for what the innate numerically relevant primitives are, and what form innate support for logic takes. If arithmetic can be derived from the resources of logic alone (with no numerical primitives), this is relevant to the question of the origin of arithmetic in ontogenesis *only* if the relevant logical resources are innate, and in a form that would support the relevant construction. If primitives with numerical content are needed as well (e.g., the principle that 1-1 correspondence guarantees cardinal equivalence, or the concepts ONE and SUCCESSOR), then one must account for how these arise in development. *TOOC* provides evidence that these numerical concepts are not part of the child's innate endowment, and that they arise only after the

bootstrapping episode in which the numeral list representation of number is constructed.

TOOC does not consider the form innate support for logic takes, and how logical resources arise in development. Indeed, I am acutely aware of this lacuna, and of its relevance to our understanding of numerical development. These questions have been the focus of research in my lab for the past four years, and will be so for the next decade at least. We do not yet have answers concerning the form innate support for logic takes. My current guess is that innate logic is largely implicit, embodied in computations, and that bootstrapping is needed before children create the logical resources needed for the mathematical construction of the integers from such primitives. After all, these constructions did not arise in mathematics until after 2000 years of development of formal logic. However, as I say below, my picture of the ontogenesis of concepts of integers would be falsified by the discovery of manifest representations with numerical content in addition to the three systems for which we already have empirical support.

Thus, I acknowledge that Fodor (2010), Leslie et al. (2007), Rey (2014), Rips et al. (2008), and others *could turn out to be* right (not that they provide a shred of evidence) that a full characterization of the manifest initial state will reveal expressive power sufficient to express arithmetic. If so, I would certainly back away from my claims about this bootstrapping episode increasing expressive power, saying that my studies concern how arithmetic capacities *actually* become manifest in ontogenesis. After all, the latter is actually my concern. I am quite certain that children do not construct arithmetic as Peano/Dedekind or Frege did, and I favor my bootstrapping story about what children actually do. But, if numerical or logical primitives are needed that themselves arise as a result of bootstrapping processes, then my claims of increases in expressive power stand.

At any rate, the actual process through which representations of integers arise is an existence proof of the possibility that bootstrapping *can* yield new primitives. The case study of the ontogenetic origin of integer representations illustrates all three major theses in *TOOC*: the existence of conceptually rich innate representations, conceptual discontinuity, and Quinian bootstrapping.

10 Core cognition of number (rich innate representational resources; *TOOC*, Chapter 4[4])

Core cognition contains two systems of representation with numerical content: parallel individuation of small sets of entities in working memory models, and analog magnitude representations of number. Analog magnitude representations were briefly sketched in section 3 above. They are analog symbols of approximate cardinal values of sets. One signature of this system of number representation is that magnitudes are compared to one another on the basis of their ratios, and thus discriminability accords with Weber's law (discriminability is better the smaller the absolute value of the quantity) and exhibits scalar variability (the standard deviation of multiple estimates of a given quantity is a linear function of the absolute value of that quantity.) Analog magnitude representations of number have been demonstrated in many animals (rats, pigeons, non-human primates) as well as in humans from neonates to adults.

Analog magnitude representations are the output of paradigmatic perceptual input analyzers, but the analog magnitude symbols for number that are produced are conceptual in the sense of having rich central conceptual roles, including the many different arithmetical computations they enter into, and the fact that they are bound to (quantify over) many types of individuals (objects, events, auditory individuals).

A second system of core cognition with numerical content, parallel individuation, consists of working memory representations of small sets of individuals (three or fewer). The symbols in this system represent individuals (e.g., a set of 3 crackers is represented CRACKER CRACKER CRACKER, probably with iconic symbols for each cracker). Unlike the analog magnitude number representation system, parallel individuation/working memory is not a dedicated number representation system, nor are there any symbols that represent cardinal values (or any other quantifiers) in these models; there are only symbols for individuals. These models are used to compute total volume and area of the individuals, and are input into spatial and

[4] The evidence for central claims in *TOOC*, along with citations of relevant literature, can be found in the chapters flagged throughout the current text.

causal representations. The numerical content in the system of parallel individuation is entirely implicit; the symbols in the models stand in 1-1 correspondence with individuals in the sets modeled. This is ensured by computations sensitive to spatiotemporal cues to numerical identity. The system must determine whether a given individual is the same one or a different one from a previously viewed individual to determine whether to add another symbol to the model. Further implicit numerical content is embodied in some of the conceptual roles these models enter into. More than one model can be entertained at any given time, and models can be compared on the basis of 1-1 correspondence to establish numerical order and equivalence. Importantly, this system of representation implicitly represents *one*. There is no explicit symbol with the content *one*, but the system updates a model of a set of one when a numerically distinct individual is added to it, yielding a model of a set of two (and ditto for sets of two and three), and the system similarly updates a model if individuals are removed from it. There is a strict upper limit to the number of individuals that can be held in working memory at any given time: 3 for infants. This set-size limit on performance contrasts with the ratio limit on performance that characterizes analog magnitude systems.

The parallel individuation system is perception-like in many ways, especially if the symbols for individuals are indeed iconic, as I suspect. Nonetheless the parallel individuation models themselves are conceptual in that they are held in a working memory system that requires attention and executive function, and enter into many further computations in support of rich central inferential processes (e.g., reasoning about the actions of agents upon objects, functional analyses, causal analyses, as well as quantitative computations).

Systems of core cognition are not the only innate resources relevant to conceptual development. *TOOC* assumes also early linguistic resources, but makes no attempt to specify their exact nature (a topic for another book). And, as commented above, the nature of logical resources available to infants and toddlers is virtually unstudied. Particularly relevant for number representations are linguistic representations that underlie the meanings of natural language quantifiers. Number marking in language (quantifiers, determiners, singular/plural morphology) requires representations of sets and individuals,

and provides explicit linguistic symbols with numerical content 'a, all, some, most, many, few…'. *TOOC* reviews evidence that before age 2 children have mastered some of the basic syntax and semantics of natural language quantifiers, and that these linguistic structures provide important early constraints on the meanings of verbal numerals, via syntactic bootstrapping.

11 Conceptual discontinuity (*TOOC*, Chapter 8)

There are two steps to establishing discontinuities in development. The first, most important, step is characterizing the nature and content of symbols in successive systems of representation: Conceptual Systems 1 and 2 (CS1 and CS2). These characterizations allow us to take the second step: namely, to state precisely how CS2 is qualitatively different from CS1. With respect to numerical content, there are three CS1s: analog magnitude representations, parallel individuation, and natural language quantification.

The substantive claims in *TOOC* are that these three systems of representation exist, have been characterized correctly, and are the *only* representational systems with numerical content manifest in infancy and the toddler years. *TOOC's* picture of number development would be falsified if evidence were to be forthcoming for innate numerical representations in addition to those described above, or different from them. Indeed, one aim of my current work on the logical resources of infants and toddlers is to search for such evidence.

CS2, the first explicit representational system that represents even a finite subset of the positive integers, is the verbal numeral list embedded in a count routine. Deployed in accordance with the counting principles articulated by Gelman and Gallistel (1978), the verbal numerals implicitly implement the successor function, at least with respect to the child's finite count list. For any numeral that represents cardinal value n, the next numeral in the list represents n + 1.

CS2 is qualitatively different from each of the CS1s because none of the CS1s has the capacity to represent any integers. The new primitives are the concepts 1, 2, 3, 4, 5, 6, 7, the concepts that underlie the meanings of verbal numerals. Parallel individuation includes no summary symbols for number at all, and has an upper limit of 3 or

4 on the size of sets it represents. The set-based quantificational machinery of natural language includes summary symbols for quantity (e.g., 'some, all') and importantly contains a symbol with content that overlaps considerably with that of 'one' (namely, the singular determiner, 'a'), but the singular determiner is not embedded within a system of arithmetical computations. Also, natural language set-based quantification has an upper limit on the set sizes that are quantified with respect to exact cardinal values (i.e., TRIAL, along with, SINGULAR and DUAL). Analog magnitude representations include summary symbols for quantity that are embedded within a system of arithmetical computations, but they represent only approximate cardinal values, and their format is analog. There is no representation of exactly 1, and therefore no representation of + 1. Analog magnitude representations cannot even resolve the distinction between 10 and 11 (or any two successive integers beyond its discrimination capacity), and so cannot express the successor function. Thus, none of the CS1s can represent 10, let alone 342,689,455.

As required by CS2's being qualitatively different from each of the CS1s that contain symbols with numerical content, it is indeed difficult to learn. American middle-class children learn to recite the count list and to carry out the count routine in response to the probe 'how many', shortly after their second birthday. They do not learn how counting represents number for another 1 ½ or 2 years. Young two-year-olds first assign a cardinal meaning to 'one', treating other numerals as equivalent plural markers that contrast in meaning with 'one'. Some 7 to 9 months later they assign cardinal meaning to 'two', but still take all other numerals to mean essentially 'some', contrasting only with 'one' and 'two'. They then work out the cardinal meaning of 'three' and then of 'four'. This protracted period of development is called the 'subset'-knower stage, for children have worked out cardinal meanings for only a subset of the numerals in their count list.

Many different tasks, which make totally different information processing demands on the child, confirm that subset-knowers differ qualitatively from children who have worked out how counting represents number. Subset-knowers cannot create sets of sizes specified by their unknown numerals, cannot estimate the cardinal values of sets outside their known numeral range, do not know what set-size

is reached if 1 individual is added to a set labeled with a numeral outside their known numeral range, and so on. Children who succeed on one of these tasks succeed on all of them. Furthermore, a child diagnosed as a 'one'-knower on one task is also a 'one'-knower on all of the others, ditto for 'two'-knowers, 'three'-knowers and 'four'-knowers. The patterns of judgments across all of these tasks suggest that parallel individuation and the set-based quantification of natural language underlie the numerical meanings subset-knowers construct for numeral words.

Also consistent with the claim of discontinuity, studies of nonverbal number representations in populations of humans who live in cultures with no count list (e.g., the Piraha: Gordon 2004; Frank et al. 2008; the Munduruku: Pica et al. 2004), and populations of humans in numerate cultures with no access to a count list (e.g., homesigners, Spaepen et al. 2011) show no evidence of any number representations other than the three CS1s.

In sum, the construction of the numeral list representation is a paradigm example of developmental discontinuity. How CS2 transcends CS1 is precisely characterized, and consistent with this analysis, CS2 is difficult to learn and not universal among humans.

12 Greater expressive power?

The above analysis makes precise the senses in which the verbal numeral list (CS2) is qualitatively different from those manifest representations with numerical content that precede it: it has a totally different format (verbal numerals embedded in a count routine), none of the CS1s with numerical content can express, even implicitly, an exact cardinal value over 4. But is the argument that the concepts for specific integers are new *primitives*, undefinable in terms of preexisting concepts using the combinatorial resources available to the child, actually correct? This argument, if correct, establishes the claim that acquiring the verbal count list representation of integers increases expressive power. As I comment in *TOOC*, this is on its face an odd conclusion. Integers are definable, after all, in terms of many different possible sets of primitives (e.g., 1 and the successor function, or the principle that 1-1 correspondence guarantees numerical equivalence plus the resources of quantificational logic).

At issue is whether logical combination underlies the transition from CS1 (core cognition of number) to CS2 (representations of verbal numerals that implicitly express the successor function). This is only possible if the capacity to represent integers is innate (e.g., if there is an innate representation of ONE and SUCCESSOR), or if integers are definable, by logical construction, from *manifest* innate primitives using *manifest* logical processes of conceptual combination. Whether acquiring integer representations increases expressive power simply is this question. Without a full characterization of the manifest combinatorial (logical) apparatus available to the child at the time the integers are constructed one cannot definitively answer the question of whether the child *could in principle* construct integer representations from innate resources, quite apart from the question of whether this is how the child *does* arrive at integer representations. But one can explore how the child actually *does* do so, and, in the remaining pages of this paper, I explain why I believe the process is *not* one of logical construction.

It's true that humans must ultimately be able to formulate concepts of integers using the explicit machinery of logic, enriched by whatever numerical concepts are necessary as well (what is actual is possible). But it is only after very long historical, and ontogenetic, developmental processes that the construction of integers in terms of logic or Peano's axioms is made. We simply do not know whether part of this process involved bootstrapping new logical representations as well as new numerical primitives.

13 A logical construction of the cardinal principle

Piantadosi et al. (2012) demonstrated that children could, in principle, construct a count list representation of the integers (at least up to 'ten') by conceptual combination alone, given the full general resources of logic (in the form of logical and set operations—if/then, set difference, plus lambda calculus, including the capacity for recursion), knowledge of the structure of the count list (its order), and four numerical primitives: the concepts SINGLETON, DOUBLETON, TRIPLETON, and QUADRUPLETON (i.e., already manifest concepts of 1, 2, 3, and 4). Piantadosi et al. appeal to the literature on learning to count in support of the claim that these numerical concepts and a

representation of the count list are manifest at the time of the induction of the counting principles, but they merely assume—without evidence—that full general resources of lambda calculus and logic are available for the generation of hypotheses about what 'one', 'two', 'three', 'four', 'five'…through 'ten' mean. They assume that children learn the meanings of the words 'one' through 'ten' from hearing words in cardinal contexts, through Bayesian enumerative induction. Thus, their model satisfies Fodor's premises 1 and 2.

The model receives input in the form of sets with 1 to 10 items paired with the appropriate verbal numeral. It learns a function, in the language of lambda calculus, that allows it to answer the question 'how many individuals?' with the correct numeral. The model's input reflects the relative frequency of verbal numerals in parental speech to children (i.e., 'one' is vastly more frequent than 'two', and so on.) Learning is constrained by limiting the combinatorial primitives that articulate hypotheses to be evaluated to those detailed above, by a preference for simpler hypotheses (i.e., shorter expressions in lambda calculus), and by a parameter that assigns a cost for recursion. After considering over 11,000 (!) different hypotheses composed from these primitives, the model learns to assign the words 'one' through 'four' to the concepts SINGLETON, DOUBLETON, TRIPLETON, and QUADRUPLETON, and also (independently) learns a recursive cardinal principle knower function that assigns the numerals 'one' through 'ten' to sets of one through ten individuals. The recursive function tests whether the set in question (S) is a singleton, and if so, answers 'one'. If not, it removes an element from S, and computes 'next' in the count list. It then applies the same singleton probe on the resultant set. If the answer is now yes, it outputs the numeral achieved by the 'next' function (i.e., 'two'.) If not, it recursively repeats this step, stepping up through the count list and down through the set until a singleton results from the set difference operation.

The model matches, qualitatively, several details of children's learning to count: children go through 'one'-, 'two'-, 'three'- and 'four'- knower stages, in that order, and depending upon the cost assigned to recursion, learn the CP-knower function after becoming 'three'-knowers or 'four'-knowers. Before the model learns the recursive CP-function, it has no way of knowing what numeral to apply to sets greater than 4, and in this sense Piantadosi et al. claim

a discontinuity in the model's knowledge of number word meanings. Thus, they claim for this model that it puts bootstrapping on a firm computational basis, as well as focusing on the logical resources actually needed for bootstrapping to succeed.

Piantadosi et al. assert that combination is the source of novelty. Therefore, in the current discourse, they are denying a genuine discontinuity. There is no change in expressive power—the manifest primitives (both numerical and logical) clearly can, in combination, express the cardinal meanings of 'one' through 'ten'. I will show why this model does not implement Quinian bootstrapping after I've discussed Quinian bootstrapping (see Rips, Asmuth and Bloomfield 2013, for an illuminating discussion). Here I simply want to acknowledge that, of course, depending upon the manifest concepts (both numerical and logical) actually available to the child, it certainly could be possible to learn the meanings of verbal numerals by constructing them from antecedently available concepts through logical combination.

The question that concerns me is how representations of integers *actually* arise in development. In what follows, I sketch a very different picture, one that does not rely on conceptual combination alone, and provide reasons to believe that this is the correct picture. My goal is to provide reasons to doubt that hypothesis formation by logical combination from primitives is the *only* source of new concepts.

14 Quinian bootstrapping

In Quinian bootstrapping episodes, mental symbols are established that correspond to newly coined or newly learned explicit symbols. The latter are initially placeholders, getting whatever meaning they have from their interrelations with other explicit symbols. As is true of all word learning, newly learned symbols must necessarily be initially interpreted in terms of concepts already available. But at the onset of a bootstrapping episode, these interpretations are only partial—the learner does not yet have any manifest concepts in terms of which he or she can formulate the concepts the symbols will come to express.

The bootstrapping process involves aligning the placeholder structure with the structure of existent systems of concepts that are

manifest in similar contexts. Both structures provide constraints, some only implicit and instantiated in the computations defined over the representations. These constraints are respected as much as possible in the course of the modeling activities, which include analogy construction. When the bootstrapping is under metaconceptual control, as is the case when it is being carried out by adult scientists, the analogical processes are explicit, and the fruitfulness of the analogies are monitored, and other modeling processes are also deployed, such as limiting case analyses, and thought experiments. Inductive inference is also often involved in bootstrapping, constrained by the actual conceptual structures in the process of being aligned.

In the case of the construction of the numeral list representation of the integers, the memorized count list is the placeholder structure. Its initial meaning is exhausted by the relations among the external symbols: they are stably ordered and applied to a set of individuals one at a time. 'One, two, three, four…' initially has no more meaning for the child than 'a, b, c, d…', if 'a, b, c, d…' were to be recited while attending to individuals one at a time.

The details of the subset-knower period suggest that the resources of parallel individuation, enriched by the machinery of linguistic set-based quantification, provide numerical meanings for the first few numerals, independently of their role in the memorized count routine. Le Corre and I (2007) proposed that the meaning of the word 'one' is represented by a mental model of a set of a single individual {i}, along with a procedure that determines that the word 'one' can be applied to any set that can be put in 1-1 correspondence with this model. Similarly 'two' is mapped onto a long term memory model of a set of two individuals {j k}, along with a procedure that determines that the word 'two' can be applied to any set that can be put in 1-1 correspondence with this model. And so on for 'three' and 'four'. This proposal requires no mental machinery not shown to be in the repertoire of infants—parallel individuation plus the capacity to compare models on the basis of 1-1 correspondence. But those representations are enriched with the long-term memory models that have the conceptual role of assigning 'one', 'two', 'three', and 'four', to sets during the subset-knower stage of acquiring meanings for verbal numerals. We suggested that enriched parallel individuation might also underlie the set-based quantificational machinery

in early number marking, making possible the singular/plural distinction, and in languages that have them, dual and trial markers. The work of the subset-knower period of numeral learning, which extends in English-learners between ages 2:0 and 3:6 or so, is the creation of the long term memory models and computations for applying them that constitute the meanings of the first numerals the child assigns numerical meaning to.

Once these meanings are in place, and the child has independently memorized the placeholder count list and the counting routine, the bootstrapping proceeds as follows: The child must register the identity between the singular, dual, trial, and quadral markers and the first four words in the count list. In the course of counting the child notes (at least implicitly) the suspicious coincidence that the numeral reached when counting a set of 'three' is also the word a 'three'-knower takes to represent the cardinal value of that set. This triggers trying to align these two independent structures. The critical analogy is between order on the list and order in a series of sets related by an additional individual. This analogy supports the induction that any two successive numerals in the child's finite count list will refer to sets such that the numeral farther in the list picks out a set that is 1 greater than that earlier in the list.

In my earliest writings I characterized the induction made by 4-year-olds as yielding the first representations of integers. Let me be clear, as *TOOC* is, when the child has built the count list representation of the first 10 or so verbal numerals, the child does not yet have general representation of integers. There are many further bootstrapping episodes along the way to a representation of integers, two of which are discussed in *TOOC*—about 6 months after becoming CP-knowers, children construct a mapping between the count list and analog magnitude representations, yielding a richer representation of the meanings of verbal numerals (Chapter 9). Shortly thereafter, children abstract an explicit concept NUMBER, and explicitly induce that there is no highest number (Hartnett and Gelman 1998). And it is not until late in elementary school or even high school that children construct a mathematical understanding of division that allows them to reanalyze integers as subset of rational numbers (Chapter 9). All of these developments are along the way to richer and richer representations of integers. But without the construction of an integer

list representation of a finite subset of integers, which provides children with new primitive concepts for specific integers beyond four (e.g., 'seven' representing exactly seven) as well as providing new representations of 'one' through 'four' (in terms of their place in a count list, rather than only in terms of enriched parallel individuation) these further bootstrapping episodes never get off the ground.

This proposal illustrates all of the components of bootstrapping processes: placeholder structures whose meaning is provided by relations among external symbols, partial interpretations in terms of available conceptual structures, modeling processes (in this case analogy), and an inductive leap.

The greater representational power of the numeral list than that of any of the systems of core cognition from which it is built derives in part from creating a new representational structure—a count list—a new conceptual role—counting, and just *using it*. Much of the developmental process involves no hypothesis testing. Just as when the child learns a new telephone number (memorizes an ordered list of digits) and learns to use it in a procedure (dial, press buttons) that results in a ring and connection to Daddy, here the child learns an ordered list and procedure for applying it to individuals as one touches them one at a time. This new structure comes to have numerical meaning through the alignment of aspects of its structure with aspects of the structure of manifest number representations. These, in turn, have been built from set-based quantification (which gives the child singular, dual, trial, and quadral markers, as well as other quantifiers), and the numerical content of parallel individuation (which is largely embodied in the computations carried out over sets represented in working memory models with one symbol for each individual in the set). The alignment of the count list with these manifest meanings is mediated, in part, by the common labels (the verbal numerals) in both structures. At the end of the bootstrapping episode, the child has created symbols that express information that previously existed only as constraints on computations. Numerical content does not come from nowhere, but the process does not consist of *defining* 'seven' by conceptual combination of primitives available to infants. 'Seven' is genuinely a new primitive, the meaning of which is provided in part by its conceptual role in a new conceptual structure.

With this characterization in hand, one can see why the Piantadosi et al. (2012) model does not implement a Quinian bootstrapping process. There are three theoretically important differences between Quinian bootstrapping and a model that formulates hypotheses at random by explicit conceptual combination from 15 primitives, one numeral at a time, and then uses Bayesian induction to evaluate them. First, although, like Piantadosi et al., I assume that children have representations with the content SINGLETON, DOUBLETON, TRIPLETON, QUADRUPLETON, before the children induces the cardinal principles, the numerical content of these representations is carried by enriched parallel individuation, and is merely implicit until the child constructs the relevant structures. On this proposal there are no manifest summary discrete symbols for these concepts. The first explicit symbols are 'one', 'two', 'three' and 'four' and their meanings are not already existing primitives SINGLETON, DOUBLETON, TRIPLETON, QUADRUPLETON. Similarly, the representations that underlie the meaning of 'seven', after the cardinal principle induction, are largely implicit in the procedures of the count routine, not explicitly defined in a language of thought. Second, the meanings of numerals in the Piantadosi model are learned entirely independently from each other. That is, children could, in principle, compose the recursive definition of numerals first, without ever going through 'one'-, 'two'-, 'three'-, and 'four'-knower stages. In Piantadosi's model, although the primitive SINGLETON plays a role in the cardinal principle function, knowing the meaning of 'one' (expressing the innate primitive SINGLETON) plays no role in learning the meanings of other numerals nor learning the cardinal principle underlying how counting expresses number. In Quinian bootstrapping, the structure created by interrelations of the newly learned words, plus their partial meanings from initial mappings to prelinguistic thought, play an essential, constitutive role in the learning process. Thirdly, and relatedly, the Quinian bootstrapping story takes seriously the question on the source of constraints on the learning process. It empirically motivates its claims of the exhaustive set of primitives with numerical content, (the three CS1s), and provides evidence for syntactic bootstrapping as an account for how the child breaks into the meanings of the first numerals. As Rips et al. (2013) point out in their illuminating discussion of the Piantadosi model, this model does not

provide an account for how the hypothesis space is conveniently limited to just the 15 numerically relevant primitives it randomly generates hypotheses from. The child has much other numerically relevant knowledge at the time of the CP induction. If that knowledge were included in the set of primitives, the hypothesis space created by random combination from the primitives would explode beyond the already entirely unrealistic 11,000 hypotheses considered and rejected by the model. If numerically irrelevant primitives are included (how does the child decide which primitives are relevant?), the problem would quickly become entirely intractable.

In sum, Quinian bootstrapping is very different from the Piantadosi logical combination model, but which model provides better insight into how children actually learn how counting represents number? Two recent animal studies clarify the nature of bootstrapping, allowing us to see that it does not involve hypothesis testing over a huge space of existing concepts, nor does it involve logical combination of primitives. These studies also increase the plausibility that young children have the computational resources to engage in Quinian bootstrapping.

15 Animal models

In *TOOC* I speculated that Quinian bootstrapping might well be a uniquely human (depending upon external explicit symbols as it does), and thus might provide part of the explanation for the uniquely human conceptual repertoire. Since then, two studies have convinced me that other animals have the capacity for Quinian bootstrapping.

15.1 Alex

The first study (Pepperberg and Carey 2012) drew on explicit numerical representations created by Alex, an African grey parrot, who had been trained by Irene Pepperberg for over 30 years. He had a vocabulary of over 200 words, including object labels, color words, relational terms such as 'same', and several other types of labels. Alex had been taught to produce the words 'three' and 'four' in response to 'how many x's' for sets of 3 and 4 respectively. During

this initial training, Alex was also shown mixed sets of objects (e.g., 4 blue balls, 5 red balls, and 3 yellow balls), and asked, for example, 'what color three,' responding 'yellow.' In other words, he was first taught to produce and comprehend 'three' and 'four' as symbols for cardinal values 3 and 4. After this training was in place, he was similarly taught to produce and comprehend 'two' and 'five' as symbols for cardinal values 2 and 5. And then 'one' and 'six' were added to his repertoire.

We do not know what non-linguistic numerical representations underlay these explicit numeral representations, because we do not know the sensitivity of Alex's analog magnitude representations or the set size limit of his parallel individual system. Analog magnitude representations themselves could have done so, or both parallel individuation and analog magnitudes could have been drawn upon. As he is a non-linguistic creature, he doesn't have the resources of set-based quantification that is part of the language acquisition device to draw upon. What the quantificational resources of non-linguistic thought are has not been studied, but Alex clearly had the capacity to selective attend to small sets and evaluate whether any given set had a cardinal value of 'one' through 'six'.

After he had a firm understanding of the cardinal meanings of 'one' through 'six', Pepperberg taught him to label plastic tokens of Arabic numerals '1, 2, 3, 4, 5' and '6', with the words 'one' through 'six' respectively. Arabic numerals were never paired with sets of individuals. The only connection between Arabic numerals and set sizes was through the common verbal numeral (e.g., 'two' for '2' and 'two' for a set of 2 individuals.) He quickly learned to produce and comprehend the verbal numeral labels for the Arabic numerals. Then with no further training, Pepperberg posed him the following question for each pair of Arabic numerals between '1' and '6': 'Which color bigger?' He was to choose, for example, between a blue '3' and a red '5', the plastic Arabic numeral tokens being the same size and the correct answer being 'red'. He succeeded at this task when first presented it; it required no further training. Not only had he not been given any positive evidence that '2' refers to a cardinal value, the only context in which he had answered questions about 'bigger' and 'smaller' previously was in with regards to physical size (i.e., 'which color bigger' of two objects that differed in size).

Please dwell on this finding. It must be that the common labels (e.g., 'two') had allowed him to connect a representation of the Arabic digits (e.g.,'2') with the cardinal values (e.g., 2), and it must be that the intrinsic order in his nonverbal representations of cardinal values allowed him to say which Arabic numeral was bigger or smaller than which others. Although Alex had never been taught a count list (and had been taught the cardinal meanings of numerals in the order 'three/four', 'two/five' and finally 'one/six'), by the time we began our study Alex could produce and comprehend the words 'one' through 'six' as labeling both cardinal values of sets and Arabic digits, and could use the intrinsic order among set sizes to order his verbal numerals.

We were thus in a position to teach Alex to label Arabic numerals '7' and '8', 'seven' (pronounced by him 'sih-none' and 'eight' respectively). This training took about a year, and during it he had no evidence that '7' or '8' were numerals. He was then taught that '6' is a smaller number than '7', which in turn is a smaller number than '8', by posing the 'which color number bigger/smaller' task, giving him feedback if he guessed wrong. This was the first evidence he had that '7' and '8' are numerals, as are '1' through '6'. It took only a few hours to train him to answer which color number bigger or which color number smaller for each of the pairs: '6/7', '6/8' and '7/8'. After he had reached criterion on this task he was probed which color number bigger and smaller for each pair of numerals '1, 2, 3, 4, 5, 6' with respect to '7' and '8', and succeeded at this task with no further training. Thus, at this point he knew that '7' and '8' are verbal numerals, labeled 'sih-none' and 'eight' respectively, and he knew that '8' is a bigger number than '1' through '7' and '7' is a bigger number than '1' through '6'. Importantly, he had never been given any information about which cardinal values 'sih-none/7' and 'eight/8' referred to.

The question of this study was whether he would make the (wildly unwarranted) induction that 'sih-none/7' expresses cardinal value 7 and 'eight/8' expresses cardinal value 8. He did. The very first time he was asked to label a set of seven objects 'how many treats?' he answered 'sih-none' and the first time he was asked to label a set of eight objects 'how many treats?' he said 'sih-none' and immediately self corrected to 'eight'. Over a two week period he was asked

to label sets of different sizes. These questions were probed by many different experimenters, only a few questions a day, intermixed with many other questions currently under study, concerning visual illusions and many other things. He performed better than chance producing both 'sih-none' and 'eight' ($p < .01$ in each case). He was also given comprehension trials, (e.g., 'what color seven' and 'what color eight', probed with 3 sets or either 6, 7, 8, 9, or 10 colored blocks), and got 11 of 12 correct ($p < .01$). Thus, Alex had inferred the cardinal meanings of 'eight' and 'seven/sih-none' from knowledge of the cardinal meanings of 'one' through 'six' and from the fact that six is a smaller number than seven and seven is a smaller number than eight.

The Piantadosi model could not possibly apply here. This learning episode did not involve hypothesis confirmation. Alex never got any feedback as to whether his answers were correct. Nor was he ever given the pairings between 'seven (sih-none)' and sets of 7 and 'eight' and sets of eight that constitute the data for the Piantadosi model. Alex *must* have made an inductive inference based on the meanings of numerals he already had constructed. Given that his knowledge of the use of numerals was exhausted by just a few procedures they entered into (answering questions about set size and numerical order, labeling cardinal values of sets and labeling Arabic numerals), and by the mappings he had already made between representations of sets, verbal and Arabic numerals, his induction was subject to strong constraints. He clearly had not searched through a vast unconstrained hypothesis space specified by logical combination of all 250 or so concepts he had that were lexicalized (or even a larger set of conceptual primitives he may manifest). As mentioned, this induction was wildly unwarranted; what he had been taught was consistent with '7' referring to any set size greater than '6' and with '8' referring to any set size greater than whatever '7' refers to. But in his 30 years of working with numerals, they had been introduced as related by +1 ('three' vs. 'four', then 'two' and 'five', and then 'one' and 'six' added to his repertoire in turn). His induction was not mathematically or logically warranted, but it was sensible, given his actual experience with numerals. So too is the child's.

Piantadosi et al. might reply that Alex may have made the leap to CP knower, having engaged in the same conceptual combination process as hypothesized by their model that children do, during the

period of learning where he was taught 'one' through 'six'. In that case, the induction he made here was that 'seven' and 'eight' were the next two numerals, in that order, in the relevant list after 'six'. This is also not possible, because Alex lacked an essential set of primitive functions for the Piantadosi model during this earlier period: namely, he did not have a count list. He was never taught a list, per se, nor never taught to count. Thus he could not form any generalizations carried by the function Next applied to a count list. He wasn't even taught the numerals in numerical order (remember he learned first 'three' and 'four', then 'two' and 'five' and finally 'one' and 'six'). It's true he explicitly knew his numerals were ordered, but that order had to be derived from the intrinsic order of cardinal values that were expressed by numerals and could not have been part of the source of the mapping between numerals and cardinal values. That order was not carried by a count routine and a memorized ordered list. Further insight into the process of learning Alex was more likely engaged in is provided by a recent study of rhesus macaques.

15.2 Rhesus macaques

Livingstone et al. (2009) taught four juvenile male rhesus macaques (1 year old at beginning of training), to choose the larger of two dot arrays, or to choose a symbol that came later in an arbitrary list. The dot arrays varied between 1 and 21 dots, and the arbitrary list of symbols was '1, 2, 3, 4, 5, 6, 7, 8, 9, X, Y, W, C, H, U, T, F, K, L, N, R'. The monkeys were trained on the dot arrays and on the symbol list on alternate days. Training in both cases involved giving the monkey a choice between two stimuli (e.g., 2 dots and 7 dots, or '2' and '7') on a touch screen. When the monkey touched one of the arrays, he was rewarded with the number of pulses of juice or water that corresponded to his choice. Thus he was always rewarded, but got bigger rewards for picking the larger dot array or the symbol later in the list. The monkeys learned to pick the stimulus that led to the larger reward with both stimuli sets, and were extremely accurate with both types of stimuli, making errors only for closely adjacent values.

There were two extremely interesting results that emerged from this study. First, with no training, the first time monkeys were given

a choice between dot arrays and symbols (e.g., 4 dots and '7'), they chose the stimulus that would lead to the larger reward. That is, they had automatically integrated the two predictors of quantity of liquid—dot arrays and discrete symbols ordered in a list). Clearly this integration had to be mediated by the fact that the dot array and discrete list tasks established a common context (same testing chamber, same dependent measure of touching one of two stimuli on a screen), and the outcomes predicted were from the same scale of quantities of liquid. Still, they had integrated them. This is the structural alignment process drawn upon in bootstrapping.

Second, when making a choice between dot arrays, the noise in choices among large sets (e.g., 19 vs. 21) was greater than that between smaller sets (e.g., 9 vs. 11 or 3 vs. 5). In fact, the choices showed scalar variability, the marker of analog magnitude values (see above). But errors in choosing values on the ordered list of discrete symbols did not increase as the list got longer. Livingstone et al. interpreted this difference as showing that the mapping from dot arrays to liquid quantity showed scalar variability, whereas the mapping from the list to hedonic value was linear. A more likely interpretation is that the mapping, during learning, reflected recognizing the relevance of each type of order (order among set sizes in analog magnitude representations of number of dots, and linear order in an arbitrary list) and inducing a rule that one should pick the stimulus later in each ordering. It's analog magnitude representations of dots that showed scalar variability, and the representations of the linear order in the list that did not. It's true that some mapping between each ordering and quantity of liquid was constructed in the process, because the two orderings were integrated. But if choosing between predicted quantities of liquid underlay each choice, both tasks should have shown scalar variability, since quantity of liquid is represented with an analog magnitude value. I suggest that the structure of an ordered list of symbols is a linear order, supported by the discriminability of each symbol from each other, and this order directly determined choice once the task was learned. This structure, after being constructed, was alignable with the intrinsic order of representations of quantity of liquid, and then with the other predictor of quantity of liquid (dot arrays). This is structurally the same as the alignment between an ordered list and analog magnitude

representations of number achieved some 6 months after children have become cardinal principle knowers.

Livingstone's rhesus macaques did not induce the cardinal meaning of a new symbol from its place in a count list, but nonetheless they exhibited several components of the extended boostrapping process that supports children's (and Alex's) doing so. They did build a representation of an ordered list (21 elements long!) and align it with a representation that was itself intrinsically ordered. Also, they automatically aligned two different ordered representations (the list, the dot arrays) which were separately aligned to quantity of liquid. Clearly, finding the structural correspondence between an ordered list and increasing magnitude (whether that magnitude is a number or a continuous variable like quantity of liquid) is a natural computation, at least for primates.

15.3 Conclusions concerning the nature of bootstrapping

As the historical examples discussed in *TOOC* make clear, bootstrapping episodes are often under metaconceptual control; the scientist is consciously engaged in exploring mappings between mathematical structures and physical/biological/psychological phenomena. But as the above examples make clear, metaconceptually explicit hypothesis testing and modeling procedures are not necessary.

I now turn to the questions of whether the representations achievable by bootstrapping should be thought of as part of a preexisting hypothesis space, or otherwise as a process of formulating and confirming hypotheses in terms of concepts that are logical constructions from primitives in a preexisting hypothesis space.

Prior to the bootstrapping processes, neither children, nor Alex, nor rhesus macaques have any representations for exact cardinal values outside of the range of parallel individuation. A representation of 341,468, or of 10, does not exist in some preexisting hypothesis space ready to become manifest. Nor is a representation of 7 constructed by conceptual combination of innate primitives. Of course the child and Alex and the rhesus macaques, must have the capacity to represent linear order, and to construct a mapping between different ordered representations, but this process does not involve constructing definitions. Some of the learning processes involved in

the extended episode of bootstrapping are certainly not hypothesis testing (e.g., memorizing the ordered list of numerals), and some are subpersonal (as Shea (2011) put it 'not explainable by content'; see also Strevens' (2009) proposal that introjection involves subpersonal processes). That is, the connection of the 'three' in the count list with the 'three' of enriched parallel individuation is most probably mediated simply by the shared label and associative machinery, just as Alex's aligning of his representations of verbal numerals, set sizes, and Arabic numerals is based first on common labels, which then supports ordering them according to the intrinsic order among cardinal values within AM and parallel individuation systems of representations. Similarly, the rhesus' aligning of an ordered list of 21 discrete symbols with set sizes from 1 to 21 depends upon shared associations with quantities of liquid. Such alignment processes are not processes of logical combination (although logical combination is involved in building the placeholder structures). Also, Alex never got any feedback regarding the pairing of 'seven' and 'eight' with cardinal values, so no hypothesis confirmation or Bayesian enumerative induction was involved. I conclude that Quinian bootstrapping yields new primitives in this case, representations of integers embedded in a count list, and is a learning mechanism that does not conform to Premises 1 and 2 of Fodor's argument.

16 Critiques of Quinian bootstrapping

Rey (2014), Fodor (2010), and Rips et al. (2013) deny Quinian Bootstrapping is a learning mechanism that can increase expressive power by creating new primitives not laying in wait. They deny that Quinian bootstrapping actually creates new primitives. It may create new concepts, but they are not primitives; they must be constructible by logical combination from others. Specific versions of the challenges include (1) analogy cannot create new representational resources, as analogies require alignable structures antecedently, (2) the induction the child makes requires an antecedent appreciation of the successor function, and (3) the bootstrapping proposal fails to confront Goodman's grue problem, the problem of constraints on induction. As I hope is already clear, I believe all of these challenges to be off the mark.

With respect to the challenge that analogy requires already available representations to be aligned, I agree. The bootstrapping process is an extended one. The new representational resource is not created at the moment of the analogy and the induction alone. By the time of the induction of the counting principles, the child has indeed created the alignable structures needed for the limited induction he/she makes, just as Alex had. In the case of the child these structures are, by hypothesis, the count list and representations of the cardinal values of the numerals 'one' through 'four' supported by enriched parallel individuation. The whole process begins with the innate numerical resources (the CS1s described above), the enrichment of parallel individuation during the subset-knower stage, and the creation of the meaningless placeholder structure. Of course one needs both structures to align them. My account of the bootstrapping process specifies the origin of each structure and shows what new arises from their alignment.

I also don't agree with the second critique, that to notice sets of two differ from sets of three by a single individual, one must already represent the successor function. All one must be able to do is subtract 2 individuals from 3 individuals, and 1 individuals from 2 individuals, computations that both parallel individuation and analog magnitude representations support. The successor function, in contrast, generates an infinite series of cardinal values, whereas the knowledge the child has is initially restricted to the relations among sets of one, two, three and four (because of the set size limit on parallel individuation and the sensitivity of analog magnitude representations being limited to 3:4 or 4:5 among young preschoolers). But of course, without the capacity to subtract 2 individuals from a set of 3 individuals, and 1 individual from a set of two individuals, yielding a single individual in each case, the child could not make the induction concerning how his or her short count list works. I do not deny this knowledge must be in place for the induction; rather I present *evidence* that it is, including *how* it is (within the system of enriched parallel individuation in the case of children's learning to count), and *evidence* that precisely that induction separates subset-knowers from cardinal principle-knowers. Again, one must consider the format and computational roles of the actual representations involved. The induction the child most probably makes is that when you add

an individual to a set for which you would reach numeral N when applying the count routine, if you count the resulting set, you will reach the next word on the count list. This is not yet the successor function, and certainly doesn't presuppose the successor function.

Turning to the heart of Rey's and Rips et al.'s criticism: that I failed to take on Goodman's new riddle of induction. Rips et al.'s extended example of a possible induction consistent with the data children have available at the time of inducing the counting principles is modular arithmetic. They ask: why do children not hypothesize that the counting sequence begins at 1 again after reaching some value (e.g., 10, in a mod 10 system). That is, why do they not consider the hypothesis that the list cycles, just as 'Monday, Tuesday, Wednesday,…Sunday, Monday…' does. Rey asks why children do not take 'two' to be a proper name for a set, or any of a myriad other hypotheses. There are, of course, an infinite number of hypotheses consistent with any finite set of data. Human inductive inference is profligate; so too, apparently, is parrot inductive inference. Accounting for the constraints on induction is everybody's problem. This paper has been an extended response to that critique. One place both writers go wrong is closely related to the view of possessed concepts as a vast hypothesis space, laying in wait to become manifest. If this were right (think Piantadosi et al.), the issue of constraints on induction would indeed be trenchant. As I have repeatedly said, any actual learning mechanism imposes constraints on what can be learned. Thus, part of the project of exploring an actual learning mechanism is studying what constraints are imposed by it, including constraints on induction. Of course children *could* learn a modular arithmetic (as adults can), but once integrated with analog magnitude representations, their actual hypotheses about meanings of numerals are constrained by the structure of the analog magnitude system (which extends open-endedly toward higher values), and constraints that the same words do not apply to discontinuous regions of it. Induction, in this case, is constrained by the only three systems of representations with numerical content (parallel individuation, analog magnitude representations, and natural language quantification) manifest at the time of learning.

One understands the constraints on the inductions made by 3-year-olds and by Alex by attending to the extremely limited

contexts in which these inductions (and most inductions) are actually drawn (think Alex and the rhesus macaques, as opposed to the model of Piantadosi et al., selecting among over 11,000 hypotheses consistent with the data it has received, where that large hypothesis space has been artificially constrained). The induction made during the hypothesized bootstrapping episode is constrained by the structures being aligned, and their very local conceptual roles. The scientific work involved in understanding episodes of Quinian bootstrapping is characterizing those structures, showing how they arise, and showing what new is achieved by aligning them.

17 A dual factor theory of bootstrapped concepts

Section 8 argued that dual factor theory straightforwardly applies to concepts that are easily acquired, for they are supported by innate conceptual roles that are never overturned (partially determining narrow content), and by innate perceptual input analyzers that guarantee a causal connection between entities in the world and the symbols that refer to them (partially determining wide content).

Chapter 13 of *TOOC* argues that dual factor theory is also needed to understand the nature of concepts that are the output of the bootstrapping episodes that underlie the origin of concepts that are hard to attain. Space does not permit a full discussion of this issue here. Briefly, newly coined concepts are ultimately mapped to antecedent ones that were supported by innate conceptual roles and innate input analyzers, and they inherit their wide content from that of those antecedent concepts. The placeholder structures in terms of which new concepts are introduced consist of interrelations among new concepts directly represented in an external language, not yet mapped to any already existing concepts that play any role in thought or refer to anything in the world. That is, they have *only* conceptual roles to provide their content. Bootstrapping proceeds by mapping these newly coined symbols to related symbols that are already interpreted. This process is often mediated by shared labels, but requires changes within the antecedently represented concepts, changes effected by aligning the two structures though modeling processes such as analogical mapping.

In *TOOC* (Chapter 13) I considered whether any of the conceptual

roles that play such an important role in this process determine the content of the final representations, given that they are all up for revision (and indeed, are revised in every episode of bootstrapping). The issue is that conceptual role has many roles to play in a full theory of concepts that do *not* specify narrow content, such as underlying inferences and being part of the sustaining mechanisms that connect concepts to their referents. The challenge to a dual factor theory is specifying which aspects of conceptual role, if any, actually determine content.

The proposal I made in *TOOC* was that the conceptual role that exhausts the meaning of the terms introduced in newly coined placeholder structures, and that constrains the structural alignment process through which these terms come to have wide content, is part of narrow content. But how can this be so, given that the relations expressed in placeholder structures are not analytic, but rather fall under the assumptions of psychological essentialism, and thus are assumed to be (and are) up for revision? The solution, I suggested, is to take seriously the relation between ancestor and descendant concepts in cases of true conceptual change (as opposed to cases of belief revision). Narrow content is that part of conceptual role that was part of the initial placeholder structure, or the aspects of conceptual role that led to changes at the level of individual concepts (differentiations, coalescences, changes in conceptual core) in the descendants of that initial placeholder structure.

18 Conclusions

As has long been recognized, a theory of concepts must include an account, at least in principle, of how it is possible that they are acquired, both over historical time and in ontogenesis. This problem has largely been ignored in the psychological literature on concepts within cognitive psychology. I have argued here that taking this problem seriously constrains our understanding of what concepts are. There are two broad routes to concept acquisition: the easy route that underlies episodes of fast mapping and the hard route that underlies conceptual discontinuities, and requires bootstrapping. The lesson that emerges from considering the two cases side by side is the crucial contribution of conceptual role in determining content. In

the easy cases, there are innate conceptual roles for the new concepts to play (NORTH in the night sky has an innate role to play in Bunting navigation; kind concepts are supported by an innate schema within the constraints of psychological essentialism). The hard cases differ from these in that there is no innate conceptual role for the new primitives, the new inferential role and the primitives that fill those roles must be co-constructed. The bootstrapping process includes constructing new placeholder structures whose symbols get meanings entirely in terms of their interrelations with each other, and this conceptual role then comes to have wide content through modeling processes that connect it to antecedently available representations. It is not a hard sell for psychologists to consider that inferential role must have a role to play in individuating concepts and specifying their content. Considerations of acquisition show both how deeply this is so, and provide suggestive evidence concerning the questions of which aspects of conceptual role might be content determining.

19 New directions

There is much work do be done, both on what I am calling the easy cases of concept acquisition and on what I am calling the hard cases. But here I want to draw attention to an urgent problem in this discourse that is virtually unstudied—specifying what form innate support for logic takes. We cannot evaluate Premise 2 of Fodor's argument without knowing this; we cannot know whether later developing concepts can be built from earlier available primitives by straightforward conceptual combination without this. One of the deepest issues in cognitive science is at stake. Many hold (e.g., Bermudez 2007; Penn et al. 2008) that non-human animals do not have a logic-like language of thought formulated over language-like representations of propositions, and many have suggested that these arise in development only upon learning natural language. Others (e.g., Braine and O'Brien 1998; Crain and Khlentzos 2010; Fodor 1975) hold that it is obvious that non-human animals have such representational capacities, and that babies could not learn language without it. Actually, it is not obvious one way or the other. It is possible that the capacity for logic-like conceptual combination may be part of the evolved capacity for human language and that it emerges in ontogen-

esis only in the course of language acquisition. More radically, it is possible that logical content is initially embodied only in computations defined on explicit representations, like the numerical content of parallel individuation, and that bootstrapping is needed to yield meanings of language-like symbols for logical connectives.

TOOC speculated that the format of representation of all core cognition systems is iconic, and provided evidence for this in the case of core cognition of number (both AM and PI representations). But systems of core cognition do not exhaust the innate representational repertoire. At the very least there are perceptual representations as well, and *perhaps* abstract representations of relations (e.g., CAUSE, SAME). It is less plausible that the format of these latter types of representations is iconic. Furthermore, it is completely unstudied whether infants have mental representations in their language of thought with the content of logical connectives, such as AND, OR, or NOT, but if there are, it is certain that their format of representation is not iconic. There is simply no research on logical symbols and reasoning schema in infancy using the productive methods of modern studies of infant cognition. There should be.

Susan Carey
Henry A. Morss Jr. and Elizabeth W. Morss Professor of Psychology
Psychology Department
Harvard University
33 Kirkland Street
Cambridge, Mass 02138
scarey@wjh.harvard.edu

References

Ahn, W., and Kim, N. S. 2000. The role of causal features in categorization: An overview. In *Psychology of Learning and Motivation, 40*, ed. by D. L. Medin. New York: Academic Press, 23-65.

Bélanger, J., and Hall, D. G. 2006. Learning proper names and count nouns: Evidence from 16- and 20-month-olds. *Journal of Cognition and Development* 7, 45-72.

Bermúdez, J. L. 2007. Thinking without words: An overview for animal ethics. *The Journal of Ethics* 11(3): 319-335.

Block, N. J. 1986. Advertisement for a semantics for psychology. In *Midwest studies in philosophy*, ed. by P. A. French. Minneapolis: University of

Minnesota, 615-678.
Braine, M. D. S., and O'Brien, D. P. (Eds.) 1998. *Mental Logic*. Mahwah, NJ: Lawrence Erlbaum Associates.
Carey, S. 1978. The child as word learner. In *Linguistic Theory and Psychological Reality*, ed. by J. Bresnan, G. Miller and M. Halle. Cambridge, MA: MIT Press, 264-293.
Carey, S. 1982. Semantic development, state of the art. In *Language Acquisition, State of the Art*, ed. by L. Gleitman and E. Wanner. Cambridge: Cambridge University Press, 347-389.
Carey, S. 1985. *Conceptual Change in Childhood*. Cambridge, MA: Bradford Books, MIT Press.
Carey, S. 2009. *The Origin of Concepts.* New York: Oxford University Press.
Crain, S. and Khlentzos, D. 2010. The logic instinct. *Mind and Language* 25(1): 30-65.
Dehaene, S. 1997. *The Number Sense.* New York: Oxford University Press.
Dretske, F. 1981. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
Emlen, S. T. 1975. The stellar-orientation system of a migratory bird. *Scientific American* 233: 102-111.
Fodor, J. A. 1975. *The Language of Thought.* Cambridge, MA: Harvard University Press.
Fodor, J. A. 1980. On the impossibility of acquiring "more powerful" structures: Fixation of belief and concept acquisition. In *Language and Learning*, ed. by M. Piatelli-Palmerini. Cambridge, MA: Harvard University Press, 142-162.
Fodor, J. A. 1998. *Concepts: Where cognitive science went wrong*. New York: Oxford University Press.
Fodor , J. A. 2008. *LOT 2: The Language of Thought Revisited.* Oxford: Oxford University Press.
Fodor, J. A. 2010, October 8. Woof, woof [Review of the book *The Origin of Concepts*, by S. Carey]. *Times Literary Supplement*, 7-8.
Frank, M. C., Everett, D. L., Fedorenko, E., and Gibson, E. (in press). Number as a cognitive technology: Evidence from Piraha language and cognition. *Cognition* 108: 819-824.
Gallistel, C. R. 1990. *The Organization of Learning.* Cambridge, MA: MIT Press.
Gallistel, C. R., Brown, A., Carey, S., Gelman, R., and Keil, F. 1991. Lessons from animal learning for the study of cognitive development. In *The Epigenesis of Mind: Essays in Biology and Cognition*, ed. by S. Carey and R. Gelman. Hillsdale, NJ: Erlbaum, 3-36.
Garcia, J., Kimeldorf, D. J., and Koelling, R. A. 1955. Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science* 122(3160): 157-8.
Gelman, R. and Gallistel, C. R. 1978. *The Child's Understanding of Number.* Cambridge, MA: Harvard University Press.
Gelman, S. A. 2003. *The Essential Child: Origins of essentialism in everyday thought.*

New York: Oxford University Press.
Gordon, P. 2004. Numerical cognition without words: Evidence from Amazonia. *Science* 306(5695): 496-499.
Hall, D. G. and Waxman, S. R. 1993. Assumptions about word meaning: Individuation and basic-level kinds. *Child Development* 64: 1550-1570.
Hartnett, P., and Gelman, R. 1998. Early understandings of numbers: Paths or barriers to the construction of new understandings? *Learning and Instruction* 8(4): 341-374.
Keil, F. C. 1989. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.
Keleman, D. and Carey, S. 2007. The essence of artifacts: Developing the design stance. In *Creations of the Mind: Theories of Artifacts and Their Representation*, ed. by E. Margolis and S. Lawrence. New York: Oxford University Press, 212-230.
Kripke, S. 1972/1980. *Naming and Necessity.* Cambridge, MA: Harvard University Press.
Leslie, A. M., Gallistel, C. R., and Gelman, R. 2007. Where integers come from. In *The Innate Mind: Foundations and Future*, ed. by P. Carruthers, S. Laurence, and S. Stich. Oxford: Oxford University Press, 109-138.
Levin, B. and Pinker, S. 1991. *Lexical and Conceptual Semantics.* Cambridge, MA: Blackwell.
Livingstone, M. S, Srihasam, K., and Morocz, I. A. 2009. The benefit of symbols: monkeys show linear, human-like, accuracy when using symbols to represent scalar value. *Animal Cognition* 13: 711-719.
Macnamara, J. 1986. *Border Dispute: The place of logic in psychology*. Cambridge, MA: MIT Press.
Margolis, E. 1998. How to acquire a concept. *Mind and Language* 13: 347-369.
Margolis, E. and Laurence, S. 2011. Learning matters: The role of learning in concept acquisition. *Mind and Language* 26: 507-539.
Miller, G. A. 1977. *Spontaneous Apprentices: Children and Language.* Seabury Press.
Miller, G. A. and Johnson-Laird, P. N. 1976. *Language and Perception.* Cambridge, UK: Cambridge University Press.
Murphy, G. 2002. *The Big Book of Concepts*. Cambridge, MA: MIT Press.
Penn, D. C., Holyoak, K. J., and Povinelli, D. J. 2008. Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences* 31(2), 109-129.
Pepperberg, I. and Carey, S. 2012. Grey Parrot number acquisition: The inference of cardinal value from ordinal position on the numeral list. *Cognition* 125: 219-232.
Piantadosi, S. T., Tenenbaum, J. B., and Goodman, N. D. 2012. Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition* 123: 199-217.
Pica, P., Lemer, C., and Izard, V. 2004. Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306(5695): 499-503.
Prasada S,. Ferenz K., and Haskell, T. 2002. Conceiving of entities as objects

and stuff. *Cognition* 83: 141-165.
Putnam, H. 1975. The meaning of meaning. In *Language, Mind, and Knowledge*, ed. by K. Gunderson. Minneapolis: University of Minnesota Press, 131-193.
Rey, G. 1983. Concepts and stereotypes. *Cognition* 15: 237-262.
Rey, G. 2014. Innate and learned: Carey, mad dog nativism, and the poverty of stimuli and analogies (yet again). *Mind and Language* 29: 109-132.
Rips, L. J., Asmuth, J., and Bloomfield, A. 2013. Can statistical learning bootstrap the integers. *Cognition* 128: 320-330.
Rips, L. J., Bloomfield, A., and Asmuth, J. 2008. From numerical concepts to concepts of number. *Behavioral and Brain Sciences* 31: 623-642.
Rips, L. J., and Hespos, S. J. 2011. Rebooting the bootstrap argument: Two puzzles for bootstrap theories of concept development. *Behavioral and Brain Sciences* 34: 145-146.
Rosch, E., Mervis, C. B., Gray, W., Johnson, D., and Boyes-Braem, P. 1976. Basic objects in natural categories. *Cognitive Psychology* 8: 382-439.
Shea, N. 2011. Acquiring a new concept is not explicable-by-content. *Behavioral and Brain Sciences* 34: 148-150.
Smith, C., Carey, S., and Wiser, M. 1985. On differentiation: a case study of the development of size, weight, and density. *Cognition* 21(3): 177-237.
Smith, E., and Medin, D. 1981. *Categories and Concepts.* Cambridge, MA: MIT Press.
Spaepen, E., Coppola, M., Spelke. E, Carey, S., and Goldin-Meadow, S. 2011. Number without a language model. *PNAS* 108(8): 3163-8.
Strevens, M. 2000. The essentialist aspect of naïve theories. *Cognition* 74: 149-175.
Strevens, M. 2009. Theoretical terms without analytic truths. *Philosophical Studies* 143: 91-100.
Weiskopf, D. 2008. The origins of concepts. *Philosophical Studies* 140: 359-384.
Wiser, M. and Carey, S. 1983. When heat and temperature were one. In *Mental Models*, ed. by D. Gentner and A. Stevens. Hillsdale, NJ: Erlbaum, 267-297.

Counterfactuals as Strict Conditionals

Andrea Iacona
University of Turin

BIBLID [0873-626X (2015) 41; pp. 165-191]

Abstract
This paper defends the thesis that counterfactuals are strict conditionals. Its purpose is to show that there is a coherent view according to which counterfactuals are strict conditionals whose antecedent is stated elliptically. Section 1 introduces the view. Section 2 outlines a reply to the main argument against the thesis that counterfactuals are strict conditionals. Section 3 compares the view with a proposal due to Åqvist, which may be regarded as its direct predecessor. Section 4 explains how the view differs from contextualist strict conditional accounts of counterfactuals. Finally, section 5 addresses the thorny issue of disjunctive antecedents.

Keywords
Counterfactuals, strict conditionals, modal logic, counterfactual fallacies, disjunctive antecedents.

1 Ellipticism

The line of thought that will be articulated in this paper rests on three basic assumptions. The first expresses a widely accepted idea about the meaning of counterfactuals. A counterfactual is a sentence 'If it were the case that *p*, then it would be the case that *q*', where '*p*' and '*q*' figure as the antecedent and the consequent. For example, the following sentence is a counterfactual:

(1) If kangaroos had no tails, they would topple over.

Its canonical formulation is 'If it were the case that kangaroos have no tails, then it would be the case that they topple over', where 'Kangaroos have no tails' is the antecedent and 'They topple over' is the consequent. The widely accepted idea is that the meaning of a counterfactual can be stated in terms of a quantification over possible

Disputatio, Vol. VII, No. 41, November 2015
Received: 24/07/2014 Revised: 22/02/2015 Accepted: 08/10/2015

worlds restricted by a relation of similarity. As Lewis puts it,

> 'If kangaroos had no tails, they would topple over' seems to me to mean something like this: in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over.[1]

More generally, if '*p*-world' stands for a world in which '*p*' is true, and 'the actual world' is used non-rigidly as an indexical expression that singles out the world of evaluation, the meaning of 'If it were the case that *p*, then it would be the case that *q*' may be stated as follows:

(M) In any *p*-world which is relevantly similar to the actual world, *q*.

The class of relevantly similar worlds may be characterized in different ways. One option, suggested by Stalnaker, it to say that there is a unique *p*-world most similar to the actual world. Another option, suggested by Lewis, is to say that there is a set of *p*-worlds most similar to the actual world. A third option, which will be adopted here, is to say that there is a set of *p*-worlds sufficiently similar to the actual world. The difference between 'most similar' and 'sufficiently similar' turns out clear in the case in which '*p*' is true in the actual world. For in that case there is only one world most similar to the actual world, namely the actual world itself, while there may be more than one world sufficiently similar to the actual world. Anyway, this difference is not essential for the present purposes. What will be assumed is simply that, on any sensible view of counterfactuals, (M) provides a correct analysis of their meaning.[2]

The second assumption is that counterfactuals are context sensitive, in that they have different truth conditions in different contexts. Suppose that the following sentences are used to describe an imaginary situation in which Caesar is in command in Korea:

(2) If Caesar had been in command, he would have used the atom bomb.

[1] Lewis 1973: 1. The idea goes back at least to Leibniz 1985: 146-147.

[2] The difference considered can be framed in terms of the principles called Centering and Weak Centering, as explained in Arlo-Costa 2007, section 3.3.

(3) If Caesar had been in command, he would have used catapults.

There is a sense in which (2) is true but (3) is false, and there is a sense in which (3) is true but (2) is false: in the first case one has in mind a modernized Caesar, while in the second one has in mind an unmodernized Caesar. This difference is plausibly described in terms of context sensitivity. In one context, we may attach more importance to similarities and differences of one kind, so that (2) turns out true, while in another context we may attach more importance to similarities and differences of another kind, so that (2) turns out false. The same goes for (3). More generally, a context can be defined as a set of parameters that includes a world *w* and a selection function *f* from sentence-world pairs to sets of worlds. For every sentence '*p*', *f* (*p*, *w*) is a set of *p*-worlds sufficiently similar to *w*, which means that *f* determines both the weights with which similarities in particular respects contribute to overall similarity between worlds and what qualifies as a sufficient level of overall similarity. Assuming that the meaning of a counterfactual is given by (M), different contexts may provide different interpretations of the expression 'relevantly similar to the actual world' which occurs in (M). This is to say that different contexts may determine different class of relevantly similar worlds.

The third assumption concerns logical form. To say that counterfactuals are strict conditionals is to say that they are sentences of the form $\Box(\alpha \supset \beta)$. In the standard semantics of modal logic, $\Box(\alpha \supset \beta)$ is true in a world *w* if and only if $\alpha \supset \beta$ is true in every world accessible from *w*, that is, in every world that satisfies the restriction associated with the sort of necessity that $\Box$ is intended to capture. What will be assumed here is that logical form is a matter of truth conditions: to say that a formula expresses the logical form of a sentence is to say that the formula provides a representation of the truth conditions of the sentence that can be employed in a formal explanation of its logical properties. The implications of this assumption turn out clear if one thinks that, given the second assumption, a principled distinction can be drawn between the meaning of a counterfactual and its truth conditions. While the meaning of a counterfactual is constant, its truth conditions may vary depending on context. So, if the formal representation of the counterfactual depends on its truth conditions, it must be sensitive to such variation. In other words, the

primary sense in which a formula can be said to express the logical form of a counterfactual is that in which it represents the counterfactual as it is understood in a given context. Obviously, this assumption is not very orthodox. Most philosophers would be inclined to say that a counterfactual has a fixed logical form which is determined by its syntactic structure or by its meaning. But the issue of what is logical form cannot be addressed here. In what follows it will simply be taken for granted that the idea that logical form is a matter of truth conditions is interesting enough to deserve consideration.

Given these three assumptions, the thesis that counterfactuals are strict conditionals may be phrased as follows: for every counterfactual 'If it were the case that *p*, then it would be the case that *q*' and every context *c*, there is a formula of the form $\Box(\alpha \supset \beta)$ which represents the truth conditions of the counterfactual as understood in *c*. More precisely, the view that will be considered entails that counterfactuals are strict conditionals whose antecedent is stated elliptically. On this view, which may be called *ellipticism*, 'If it were the case that *p*, then it would be the case that *q*', as uttered in *c*, is properly phrased as 'Necessarily, if *p* and things are relevantly like in the actual world, then *q*', where the content of 'things are relevantly like in the actual world' is determined by *c*. Therefore, its logical form is $\Box(\alpha \supset \beta)$, where α stands for '*p* and things are relevantly like in the actual world' as understood in *c* and β stands for '*q*'. In other words, α delimitates the set of worlds that the selection function of *c* assigns to *p* relative to the world of *c*. So the counterfactual can be represented as a strict conditional whose antecedent has two parts: one is explicit, '*p*', the other is implicit, 'things are relevantly like in the actual world'.

According to ellipticism, the fact that a counterfactual may have different truth conditions in different contexts is representable at the formal level. Consider (2), and suppose that *c* and c' are two contexts which differ in the way explained above. (2) is properly phrased as 'Necessarily, if Caesar is in command and things are relevantly like in the actual world, then he uses the atom bomb', where 'things are relevantly like in the actual world' has different contents in *c* and c'. Therefore, distinct formulas may be assigned to (2) relative to *c* and c'. That is, if (2) is represented as $\Box(\alpha \supset \beta)$ relative to *c*, then it may be represented as $\Box(\gamma \supset \beta)$ relative to c': α stands for 'things are

relevantly like in the actual world' as understood in *c*, while γ stands for 'things are relevantly like in the actual world' as understood in c'. This is consistent with a general principle about formalization that is usually taken for granted, namely, that sentences with different truth conditions must be represented by distinct formulas, that is, formulas that can have different truth values in the same model.

One way to see how the principle applies is to think about the difference between a counterfactual 'If it were the case that *p*, then it would be the case that *q*' and an overt strict conditional 'Necessarily, if *p* then *q*'. Consider the following sentence:

(4) Necessarily, if kangaroos have no tails, then they topple over.

(1) and (4) have different truth conditions. For (4) means that kangaroos topple over in any possible world in which they have no tails. So if (1) and (4) were represented by the same formula, the difference between them would not be captured at the formal level. A straightforward way to draw the distinction is to assign different formulas to (1) and (4), that is, $\Box(\alpha \supset \beta)$ and $\Box(\gamma \supset \beta)$, where α stands for 'Kangaroos have no tails and things are relevantly like in the actual world' and γ stands simply for 'Kangaroos have no tails'. This method of formalization implies that counterfactuals are covert strict conditionals. They differ from overt strict conditionals, whose antecedent is stated explicitly.

Note that, since counterfactuals and overt strict conditionals are represented by the same kind of formula, there is a clear sense in which they have the same logical form. The thesis that counterfactuals are strict conditionals, as understood here, is not intended to provide an analysis of the meaning of counterfactuals in terms of $\Box$ and $\supset$. Counterfactuals exhibit distinctive semantic features that make them differ from other conditionals, and presumably there is no formula in the language of modal logic—$\Box(\alpha \supset \beta)$ or any other—such that having a logical form expressed by that formula is both necessary and sufficient for having those features. Nonetheless, it may be claimed that counterfactuals are sentences of the form $\Box(\alpha \supset \beta)$ in virtue of those features.

Ellipticism is essentially a view about the logical form of counterfactuals. Its main point concerns the formal representation of counterfactuals, rather than the analysis of their meaning. To illustrate

this feature of ellipticism, consider the Stalnaker-Lewis view, that is, the shared fragment of the theories of counterfactuals defended by Stalnaker and Lewis. Ellipticism and the Stalnaker-Lewis view converge at the conceptual level, as they both rest on the idea that the meaning of a counterfactual is expressed by (M). More precisely, in both cases a counterfactual 'If it were the case that *p*, then it would be the case that *q*' can be evaluated as true or false relative to a context defined in the way considered, provided that the selection function is appropriately specified. The key difference between the two views concerns the formal representation of counterfactuals. Stalnaker and Lewis think that a special symbol, say $>$, should be employed to capture the meaning of 'If it were the case that..., then it would be the case that...', hence that a special formal system that encompasses that symbol must be tailored to counterfactuals. According to ellipticism, instead, no logical adjustment of that kind is required. The only symbols needed are $\Box$ and $\supset$, with their familar semantics.[3]

2 Counterfactual fallacies

The main argument provided so far against the thesis that counterfactuals are strict conditionals is due to Stalnaker and Lewis. According to Stalnaker and Lewis, the thesis may appear tenable if one looks at a single counterfactual, but it proves inadequate if one reflects on sets of counterfactuals and the logical relations they involve. For at least three basic inference rules that hold for strict conditionals do not hold for counterfactuals, that is, there are at least three distinctive "counterfactual fallacies". The first is the fallacy of *strengthening the antecedent*. Consider the following argument:

A1 (5) If Otto had come, it would have been a lively party.
$\therefore$ (6) If Otto and Anna had come, it would have been a lively party.

Imagine that Otto is a very cheerful person, but that he just broke up

[3] In this respect, ellipticism differs from any attempt to define counterfactuals in terms of some characteristic modal operator analogous to $\Box$, such as Burks 1951.

with Anna after six months of endless rows. In such a situation (5) may be true even though (6) is false. In other words, (5) is consistent with

(7) If Otto and Anna had come, it would have been a dreary party.

Therefore, A1 is invalid. But the following argument form is valid:

S1 $\Box(\alpha \supset \beta)$
$\therefore$ $\Box((\alpha \wedge \gamma) \supset \beta)$

For if β is true in all accessible α-worlds, *a fortiori* it will be true in all accessible α-worlds in which γ is true. So A1 cannot instantiate S1.[4]

The second is the fallacy of *transitivity*. Consider the following argument:

A2 (8) If Otto had gone to the party, Anna would have gone.
(9) If Anna had gone to the party, Waldo would have gone.
$\therefore$ (10) If Otto had gone to the party, Waldo would have gone.

Imagine that Waldo fancies Anna, although he never runs the risk of meeting his successful rival Otto. Imagine also that Otto was locked up at the time of the party, so that his going to the party is a remote possibility, but that Anna almost did go, as she hoped to meet him. In such a situation (8) and (9) may be true even though (10) is false. Therefore, A2 is invalid. However, the following argument form is valid:

S2 $\Box(\alpha \supset \beta)$
$\Box(\beta \supset \gamma)$
$\therefore$ $\Box(\alpha \supset \gamma)$

For if all accessible α-worlds are β-worlds and all accessible β-worlds are γ-worlds, then all accessible α-worlds are γ-worlds. So A2 cannot instantiate S2.[5]

[4] Stalnaker 1991: 38, Lewis 1973: 10-13 and 31. The sequence formed by (3) and (5) is called a "Sobel sequence", from Lewis 1973:10 fn.

[5] Stalnaker 1991: 38, Lewis 1973: 32-33. Note that S2 entails S1, as it is easily

The third is the fallacy of *contraposition*. Consider the following argument:

A3 (11) If Otto had gone to the party, Anna would have gone.
∴ (12) If Anna had not gone, Otto would not have gone.

Imagine that Otto wanted to go to the party but stayed away just to avoid Anna, while Anna would definitely have gone if Otto had been around. In such a situation (11) may be true even though (12) is false. Therefore, A3 is invalid. However, the following argument form, S3, is valid:

S3 $\Box(\alpha \supset \beta)$
∴ $\Box(\sim \beta \supset \sim \alpha)$

For $\alpha \supset \beta$ and $\sim \beta \supset \sim \alpha$ have the same truth value in every accessible world. So A3 cannot instantiate S3.[6]

The Stalnaker-Lewis argument may be summarized as follows. Suppose that counterfactuals are strict conditionals. Then A1-A3 instantiate S1-S3. But A1-A3 are invalid arguments, while S1-S3 are valid argument forms. Therefore, counterfactuals are not strict conditionals. Ellipticism provides a reason to reject this argument, as it undermines the assumption that if counterfactuals are strict conditionals then A1-A3 instantiate S1-S3. Consider A1. If (5) is represented as $\Box(\alpha \supset \beta)$, then α does not stand for 'Otto has come' but for 'Otto has come and things are relevantly like in the actual world'. So (6) cannot be represented as $\Box((\alpha \wedge \gamma) \supset \beta)$. For its whole antecedent is 'Otto and Anna have come and things are relevantly like in the actual world', in which neither conjunct amounts to α. Therefore, the argument form instantiated by A1 is not S1 but the following:

S4 $\Box(\alpha \supset \beta)$
∴ $\Box(\gamma \supset \beta)$

seen if α is replaced with $\alpha \wedge \beta$. So the failure of S1 alone suffices to discard S2.

[6] Stalnaker 1991: 39, Lewis 1973: 35.

Consider A2. If (8) is represented as $\Box(\alpha \supset \beta)$, then (9) cannot be represented as $\Box(\beta \supset \gamma)$ but rather as $\Box(\gamma \supset \delta)$. Therefore, the argument form instantiated by A2 is not S2 but the following:

S5 $\Box(\alpha \supset \beta)$
$\Box(\gamma \supset \delta)$
$\therefore$ $\Box(\alpha \supset \delta)$

Finally, consider A3. If (11) is represented as $\Box(\alpha \supset \beta)$, the antecedent of the formula that represents (12) cannot be $\sim\beta$ but a different formula γ. Similarly, its consequent cannot be $\sim\alpha$ but a different formula δ that stands for 'Otto has not gone'. Therefore, the argument form instantiated by A3 is not S3 but the following:

S6 $\Box(\alpha \supset \beta)$
$\therefore$ $\Box(\gamma \supset \delta)$

Since S4-S6 are invalid forms, the invalidity of A1-A3 is easily explained.

Two final remarks. The first concerns the assumption that A1-A3 are invalid. The ellipticist reply to the Stalnaker-Lewis argument grants this assumption: its point is that, even though A1-A3 are invalid, their invalidity is no evidence against the thesis that counterfactuals are strict conditionals. However, it is important to note that here it is not essential to assume that validity is a property of arguments, at least if arguments are understood in the usual way as sets of sentences. Perhaps the most plausible thing to say, given the context-sensitivity of counterfactuals, is that validity is a property of "interpreted" arguments, where an interpretation of an argument is an assignment of contexts to its sentences. If validity is so understood, the assumption to be granted is rather that A1-A3 are invalid in the intended interpretation. This by itself does not rule out the possibility that there are interpretations in which A1-A3 or other structurally similar arguments are valid. For example, Lowe suggests that there are non-fallacious cases of transitivity:

> Suppose that two people are discussing the influence of upbringing and social background upon a person's political convictions, and one

> of them, *X*, takes Margaret Thatcher as an example of someone who, though a firm supporter of the capitalist free market economy, might have had a quite different attitude towards it. *X* reasons as follows:[...] If Margaret Thatcher had been born and brought up in the Soviet Union, she would have had communist sympathies; and if she had had communist sympathies, she would have been opposed to the capitalist free market economy. So, if she had been born and brought up in the Soviet Union, she would have been opposed to the capitalist free market economy. *X*'s reasoning seems unexceptionable to the point of appearing almost banal.[7]

Lowe's example seems to show that an argument structurally similar to A2 can be valid in some interpretation. To see how this case differs from that of A2, it suffices to think that in this case, unlike in that of A2, the possible circumstances that one has in mind when one asserts the first premise also sustain the second premise. To put it another way, the possible circumstances that one has in mind when one asserts the first premise also justify strenghtening its consequent by adding the antecedent of the second premise as a conjunct. The following sentences seems equally assertable in the situation described:

(13) If Margaret Thatcher had been born and brought up in the Soviet Union, she would have had communist sympathies.

(14) If Margaret Thatcher had been born and brought up in the Soviet Union, she would have had communist sympathies and she had been opposed to the capitalist free market economy.

Obviously, this leaves open the question of how a principled distinction can be drawn between apparently fallacious cases and apparently non-fallacious cases. However, all that matters here is that ellipticism is consistent with the possibility that an argument structurally similar to A2 is valid in some interpretation, provided that no such argument is classified as a case of transitivity. S5 is an invalid argument form. But an invalid argument form can have valid instances.

The second remark concerns A3, which illustrates the difference between ellipticism and the Stalnaker-Lewis view explained in section 1. Stalnaker and Lewis claim that contraposition does not hold

[7] Lowe 1990: 80.

for counterfactuals: A3 is an invalid argument from $\alpha > \beta$ to $\sim\beta > \sim\alpha$. Nonetheless, they maintain that *modus tollens* is a valid argument form. That is, from $\alpha > \beta$ and $\sim\beta$ we can infer $\sim\alpha$. This is why Lewis provides a non-orthodox justification of *modus tollens* that does not appeal to contraposition. He says that *modus tollens* is acceptable because from $\alpha > \beta$ we can infer $\alpha \supset \beta$, and from the latter we can infer $\sim\beta \supset \sim\alpha$ (contraposition does hold for $\supset$), so that $\sim\alpha$ follows by *modus ponens*. Even granting Lewis' justification, however, it is hard to resist our inclination to regard contraposition and *modus tollens* as different expressions of the same principle, and so to think that they should either stand or fall together. In a standard deduction system of modal logic, this inclination is vindicated by the rule of conditional proof: if one can derive β from α and auxiliary premises, then one can derive $\alpha \supset \beta$ from the auxiliary premises alone. This means that if $\sim\alpha$ follows from $\sim\beta$ and $\alpha \supset \beta$, then $\sim\beta \supset \sim\alpha$ follows from $\alpha \supset \beta$. Ellipticism,unlike the Stalnaker-Lewis view, implies no separation between contraposition and *modus tollens*. According to ellipticism, both contraposition and *modus tollens* hold: arguments such as A3 simply have little to do with them.[8]

3 The selection operator view

In the past, some attempts have been made to provide an analysis of counterfactuals that employs the expressive resources of modal logic. Ellipticism bears close resemblance to one of them, due to Åqvist. According to Åqvist, counterfactuals can formally be represented in a modal language that contains an operator $*$ whose semantics is given in terms of a selection function *f* that assigns sets of worlds to formulas. That is, $*\alpha$ is true in all and only the worlds that belong to $f(\alpha)$, where $f(\alpha)$ is understood as the set of α-worlds most similar to the actual world. In such a language, the logical form of 'If it were the case that *p*, then it would be the case that *q*' may be expressed as $\Box(*\alpha \supset \beta)$, where α stands for '*p*' and β stands for '*q*'. So it turns out that the counterfactual is true if and only β is true in all α-worlds

[8] Stalnaker 1991: 39, Lewis 1973: 36. This section is drawn from Iacona 2011.

most similar to the actual world.[9]

Ellipticism has much in common with this view, which may be called *the selection operator view*. First, the central claim of both views is that the logical form of 'If it were the case that *p*, then it would be the case that *q*' is expressed by a strict conditional whose antecedent does not stand for '*p*' but for a stronger condition that is implicit in the counterfactual. Second, both views assume that the implicit condition involves a similarity constraint in accordance with (M). Third, both views grant that the understanding of the similarity constraint may be irreducibly indexical, in that they do not require that the implicit condition amounts to a set of sentences whose conjunction provides a complete characterization of the set of worlds that satisfy the constraint.[10]

The obvious difference between ellipticism and the selection operator view is that ellipticism represents the whole antecedent of 'If it were the case that *p*, then it would be the case that *q*' as α, so it requires no special symbol to be added to the language of modal logic. Given this difference, it is natural to wonder whether there are reasons to think that one of the two views is better than the other. One might be tempted to say that the selection operator view is preferable in that a representation that involves the operator $*$ displays a relation between the explicit part and the implicit part of the antecedent that a simple formula is unable to capture. But this temptation must be resisted. As it will be suggested, ellipticism is preferable in another respect, because it provides a neat account of some fundamental modal properties of counterfactuals that trouble the selection operator view. Therefore, all things considered it is not obvious that the selection operator view is better than ellipticism.

The selection operator view comes in at least two versions: one is the original version set out by Åqvist, the other is an amended version sketched by Lewis. Let us start with the original version. The semantics for $*$ provided by Åqvist is rigidly centred on the actual world. Every model includes a distinguished world w_0, and the function f is defined in terms of w_0: the set that f assigns to each formula

[9] Åqvist 1973: 2-3.

[10] In this respect, both ellipticism and the selection operator view differ from what Lewis calls "the metalinguistic theory", see Lewis 1973: 66-67.

α is understood as the set of α-worlds most similar to w_0. However, as Lewis has argued, such a semantics is unable to account for the apparent contingency of some counterfactuals. Consider the following:

(15)If I had looked in my pocket, I would have found a coin.

Since I actually have a coin in my pocket, (15) is true in the actual world. But in a world in which my pocket is empty, (15) is false. This fact cannot be explained if $*$ is interpreted in the way considered. Certainly, if α stands for 'I looked in my pocket', which is the explicit part of the antecedent, then $*\alpha$ stands for 'I looked in my pocket and things are relevantly like in w_0', so the actual truth of (15) is explained in terms of the truth of $\Box(*\alpha \supset \beta)$. But no explanation can be provided of the falsity of (15) in a world w_1 in which my pocket is empty. As Lewis observes, this is a serious limitation. Even if we are ultimately interested in the actual world, we must consider the truth values of counterfactuals at other worlds to obtain the actual truth values of sentences in which counterfactuals are embedded inside other counterfactuals. Consider the following:

(16)If I had looked in my pocket, I would have found a coin, but if my pocket were empty, it would not be the case that if I had looked in my pocket, I would have found a coin.

The actual truth of (16) can be explained only if the semantics makes room for the possibility that different sets of worlds are associated to the same antecedent.[11]

The amended version of the selection operator view is intended to make room for that possibility. As Lewis has explained, the view can be modified in order to account for the contingency of counterfactuals such as (15). His suggestion is that f is replaced by a two-argument function f' that assigns sets of worlds to formula-world pairs, and that a three-place truth relation for $*\alpha$ is defined as follows: $*\alpha$ is true in a world w with reference to a world w' if and only if w belongs to $f'(\alpha,w')$, that is, if and only if w is one of the α-worlds most similar to w'. The three-place truth relation is then generalized to any formula by stipulating that the formula is true in w with reference to w' if and only if it is true in w. So it turns out that $\Box(*\alpha \supset \beta)$ is true in

[11] Lewis 1973: 61-62.

w with reference to w' if and only if $*\alpha \supset \beta$ is true with reference to w' in every world accessible from w, that is, if and only if β is true in every world in $f'(\alpha,w')$ accessible from w. This way we get that there can be two worlds w_0 and w_1 such that $\Box(*\alpha \supset \beta)$ is true in w_0 with reference to w_0, while it is false in w_1 with reference to w_1.[12]

However, some important questions remain open. In the first place, it is not entirely clear how to make sense of the assumption that the intension of $*\alpha$—the set of worlds in which $*\alpha$ is true—varies as a function of the world of reference. This assumption is intended to guarantee that the implicit part of the antecedent of a counterfactual can express different conditions relative to different worlds. To illustrate, let the meaning of (15) be stated as follows:

(15_M) In any world in which I looked in my pocket, and in which things are relevantly like in the actual world, I found a coin.

A straightforward way to explain why (15) is true in the actual world w_0 but false in a world w_1 where my pocket is empty is to say that the implicit past of its antecedent expresses different conditions relative to w_0 and to w_1. This means that 'the actual world' in (15_M) refers to w_0 in the first case and to w_1 in the second. That is,

($15w_0$) Necessarily, if I looked in my pocket and things are relevantly like in w_0, then I found a coin.

($15w_1$) Necessarily, if I looked in my pocket and things are relevantly like in w_1, then I found a coin.

The point, however, is that if the antecedent of a counterfactual expresses different conditions relative to different worlds, it is not clear why a single formula with variable intension should be used to represent those conditions. Note that the variation of intension at issue is not the familiar variation of intension due to a difference of model, but a variation of intension that occurs *within a model*. Of course, any standard semantics for a formal language allows that the same formula has different intensions in different models, given that models are normally understood as interpretations of the language. For example, the same formula can be read as 'Snow is white' in one case

[12] Lewis 1973: 62-63.

and as 'Grass is green' in another case. But the variation of intension involved here is of a different kind, because it implies something like saying that the same formula, in the same model, is true in w because it means 'Snow is white' relative to w and false in w' because it means 'Grass is green' relative to w'. This is quite an odd thing to say.

In the second place, Lewis does not explain how exactly the amended version of the selection operator view accounts for the actual truth of (16). In (16), two occurrences of the same antecedent hide different implicit conditions. The first condition concerns the actual world, while the second concerns a world in which my pocket is empty. Since both conditions are represented by $*\alpha$, (16) must be formalized as a complex sentence in which $*\alpha$ occurs twice. Presumably, the sentence must be such that its truth in w_0 with reference to w_0 depends on the truth of the second occurrence of $*\alpha$ having a given intension relative to a different word w_1. However, the details of the account are still missing. Unless a formal semantics is spelled out with the due accuracy, it is hard to judge whether the problem has been solved.

In the third place, it may rightfully be asked whether the amended version of the selection operator view substantially preserves the thesis that counterfactuals are strict conditionals. If a counterfactual is represented by a formula $\Box(*\alpha \supset \beta)$ that can be true in a world w with reference to w but false in a world w' with reference to w', then it can be treated as an ordinary contingent sentence. As Lewis suggests, if an operator $\dagger$ is so defined that $\dagger\alpha$ is true in w if and only if α is true in w with reference to w, we get that the counterfactual amounts to a contingent sentence of the form $\dagger\Box(*\alpha \supset \beta)$. But a sentence of that form, it might be argued, is *not* a strict conditional. Independently of the presence of $\dagger$, the obvious difference is that a sentence of that form can be contingent, while a strict conditional must be necessary if it is true. For a strict conditional is a sentence of the form $\Box\alpha$, and $\Box\alpha$ entails $\Box\Box\alpha$. Or at least, this holds on the assumption that necessity obeys S5 or similar systems.[13]

From the foregoing considerations it turns out that it is not clear

[13] The operator $\dagger$ is introduced in Lewis 1973: 63. The point that strict conditionals must be necessary if they are true is made in Sider 2010: 200, where it is used against the thesis that counterfactuals are strict conditionals.

how the selection operator view can cope with the issue of contingency. Ellipticism differs in this respect, in that it makes room for a distinction that explains how the apparent contingency of some counterfactuals squares with the thesis that counterfactuals are strict conditionals. According to ellipticism, there is a sense in which a counterfactual may be used to say the same thing in different worlds, and there is a sense in which it may be used to say different things in different worlds. The first is that in which the counterfactual has a meaning that does not vary from world to world, the meaning expressed by (M). The second is that in which the counterfactual has different truth conditions in different worlds, given that the reference of the expression 'the actual world' which occurs in (M) varies from world to world. Since a context is a set of parameters which includes a world, this is to say that the reference of that expression may vary from context to context. In the first sense, the counterfactual may be contingent. In the second, it is necessary if true. For example, (15) may be used to say something true relative to w_0 and something false relative to w_1. But what is said relative to w_0, that is, $(15w_0)$, is necessary if true. Similarly, what is said relative to w_1, that is, $(15w_1)$, is impossible if false.

The contrast between ellipticism and the selection operator view emerges clearly if one considers the three questions raised above. In the first place, ellipticism does not need to assume that there are special formulas whose intension can vary within a model. In order to account for the fact that the antecedent of a counterfactual can express different conditions relative to different worlds, it is simply assumed that different conditions require different formulas. For example, $(15w_0)$ and $(15w_1)$ are represented as $\Box(\alpha \supset \beta)$ and $\Box(\gamma \supset \beta)$. Since 'things are relevantly like in the actual world' expresses different conditions relative to w_0 and w_1, different formulas α and γ are used to represent those conditions. Therefore, the fact that (15) is true in w_0 but false in w_1 can be explained in terms of the platitude that different formulas may have different truth values.

In the second place, ellipticism can explain the actual truth of (16) in the same way. Since (16) contains two occurrences of 'I looked in my pocket' which are associated to different sets of worlds, these two occurrences are represented by different formulas, say α and γ. So the formalization of (16) does not require two occurences

of α. More generally, the method of formalization suggested is able to account for the semantic variation that affects the implicit part of the antecedent of counterfactuals, not only when this variation depends on the intended relations of similarity between worlds, but also when it depends on the world of utterance. This feature may pass unnoticed if one restricts attention to the truth values of counterfactuals in the actual world, but it becomes manifest when one considers the truth values that counterfactuals have in possible worlds different from ours.

In the third place, ellipticism definitely preserves the thesis that counterfactuals are strict conditionals. The understanding of the thesis suggested implies that strict conditionals are necessary if true. We saw that, although a counterfactual may be contingent in one sense, it may not in another sense. Since its formal representation as a strict conditional concerns the second sense, on the assumption that logical form is a matter of truth conditions, it turns out that the strict conditional must be necessary if true. Therefore, if a strict conditional is true in a world, its necessitation is also true in that world, in accordance with the S5 entailment from $\Box\alpha$ to $\Box\Box\alpha$. For example, ($15w_0$) is formally represented as $\Box(\alpha \supset \beta)$, where α stands for 'I looked in my pocket and things are relevantly like in w_0'. So $\Box(\alpha \supset \beta)$ expresses something about w_0 that is true in every world. This is why $\Box\Box(\alpha \supset \beta)$ is also true in w_0.

To sum up, the opposing inclinations towards contingency and necessity that emerge from the discussion of the selection operator view can be explained in terms of the distinction between meaning and truth conditions. Of course, one might still object that this distinction does not suffice, and insist that the intuition of contingency implies that the truth conditions of counterfactuals are themselves contingent. But nothing can be done to move such unsatisfied objector. First of all, the intuition of contingency, if there is such a thing, can hardly be so definite as to entail that it is not enough to say that the same sentence, with the same meaning, can be true in a world but false in another world. In the second place, it is reasonable to expect that a distinction along the lines suggested is the best that a strict conditional analysis can offer with respect to the issue of contingency. For a strict conditional analysis cannot rule out necessity altogether. As noted above, it would make little sense to claim that

counterfactuals are sentences of the form $\Box(\alpha \supset \beta)$ but deny that $\Box$ obeys S5 or similar systems. So the unsatisfied objector must think that in principle no strict conditional analysis can work.

4 Contextualism

Although it is generally taken for granted that counterfactuals are context sensitive, it is not entirely obvious to what extent they are context sensitive. One major point of controversy concerns the role of the antecedent in the determination of context. On the one hand, anyone agrees that the fact that different sets of worlds can be assigned to the same antecedent is correctly described in terms of context sensitivity. For example, it is plausible to say that (2) and (3) are true in different contexts, in that they are true relative to different ways of delimiting the class of relevantly similar worlds in which Caesar was in command. On the other hand, there is no equally shared account of the fact that, normally, different sets of worlds are assigned to different antecedents. For example, when (5) and (6) are evaluated respectively as true and false, the set of worlds that count as relevantly similar in the first case, those in which Otto has come to the party, differs from the set of worlds that count as relevantly similar in the second, those in which Otto and Anna have come to the party. The question, however, is whether this difference amounts to a difference of context: one option is to say that it does, the other is to say that it does not.[14]

The thesis that counterfactuals are strict conditionals is often associated with the first option. According to a line of thought that has been amply debated in the last few years, counterfactuals are highly context sensitive strict conditionals, in that their strictness varies as a function of their antecedent. Thus, (5) and (6) are strict conditionals assessed respectively as true and false in different contexts c and c', that is, they involve different accessibility relations. The intended reading of (5) is that every accessible$_c$ world in which Otto has come to the party is a world in which the party is lively. The intended reading of (6), instead, is that every accessible$_{c'}$ world in which Otto and

[14] This question is explicitly addressed in Brogard and Salerno 2008, and in Cross 2011.

Anna have come to the party is a world in which the party is lively. This means that A1 involves a context-shift, and the same goes for A2 and A3.[15]

If a strict conditional analysis of this kind is called *contextualism*, ellipticism differs from contextualism. Ellipticism rests on the assumption that counterfactuals are context sensitive in the less controversial sense, and contemplates no reason to think that they are context sensitive in the more controversial sense. As it turns out from section 1, a context may be defined in terms of a selection function. Consider two counterfactuals 'If it were the case that *p*, then it would be the case that *q*' and 'If it were the case that *r*, then it would be the case that *q*', and let *c* be a context which includes a world *w* and a selection function *f*. Since '*p*' and '*r*' are different sentences, $f(p,w)$ may differ from $f(r,w)$. But the context does not change, for *f* is the same function. This turns out clear if the two counterfactuals are represented as $\Box(\alpha \supset \beta)$ and $\Box(\gamma \supset \beta)$, where α expresses an inclusion condition for $f(p,w)$ and γ expresses an inclusion condition for $f(r,w)$. For such representation requires no variation in the accessibility relation: $\Box$ expresses unrestricted necessity in both cases. In substance, ellipticism is a non-contextualist strict conditional analysis of counterfactuals. Its mere existence shows that the issue of how the context sensitivity of counterfactuals is to be understood must not be confused with the question of whether counterfactuals are strict conditionals.

Although an examination of the arguments that may be invoked to justify contextualism goes beyond the scope of this paper, at least one issue deserves attention. Contextualism, just like any strict conditional analysis, must provide a reply to the Stalnaker-Lewis argument. For that argument questions the thesis that counterfactuals are strict conditionals. However, it seems that none of the replies available to the advocates of contextualism is preferable to that outlined in section 2.

[15] The supposition that the counterfactuals in a Sobel sequence—hence in A1—are strict conditionals that involve different contexts, initially dismissed in Lewis 1973, is developed in Von Fintel 2001 and in Gillies 2007. Similarly, Warmbrod (1981), Lowe (1990) and Lowe (1995) suggest that arguments such as A2 are affected by context-shifts, and Tichý (1984) says the same of arguments such as A3.

The Stalnaker-Lewis argument is a reductio: the thesis that counterfactuals are strict conditionals is taken to entail the absurd consequence that A1-A3 instantiate S1-S3. Therefore, in order to reject the argument, it must be contended either that the thesis does not have the alleged consequence, or that the alleged consequence is not absurd. Perhaps the most natural option for the advocates of contextualism is the second. They might draw inspiration from Kaplan's treatment of arguments containing indexicals, and reply that it is wrong to assume that A1-A3 are invalid, for in order to assess A1-A3, the context must be held fixed. According to Kaplan, an argument containing indexicals is valid if and only if, for any context, if the premises are true in that context, the conclusion must be true in that context. For example, 'She is there, so she is there' turns out valid on Kaplan's definition, because it can't be the case that a context makes 'She is there' true and false at the same time. A similar treatment may be applied to A1-A3: since validity amounts to truth preservation in any context, the fact that A1-A3 have true premises and false conclusion in the intended interpretation does not show that they are invalid, given that their intended interpretation involves context-shifts.[16]

This reply is not entirely satisfactory. If one assumes, following Kaplan, that validity is a property of arguments, and claim that A1-A3 are valid, despite the fact that their intended interpretation involves context-shifts, one has a straightforward account of the relation between A1-A3 and S1-S3: A1-A3 are valid in that they instantiate S1-S3. The obvious drawback of this reply, however, is that it clashes with the apparent invalidity of A1-A3 in the intended interpretation. Kaplan's definition leaves unexplained the fact that A1-A3 can be used in such a way that their premises are true and their conclusion is false, just as it leaves unexplained the fact that 'She is there, so she is there' can be used in such a way that its premise is true and its conclusion is false. If an argument is valid, one may be tempted to say, how can it be the case that its premises are true and its conclusion false? As it has been argued against Kaplan, a definition

[16] Kaplan's definition is suggested in Kaplan 1989. A reasoning along the lines considered is offered in Lowe 1990 and in Brogaard and Salerno 2008, although it is not accompanied by a strict conditional analysis of counterfactuals.

of validity that holds for arguments containing context sensitive expressions should take into account non-univocal interpretations of their premises and conclusions, that is, interpretations which involve context-shifts.[17]

A different way to question the assumption that A1-A3 are invalid is to assume that validity is a property of interpreted arguments, and claim that, although A1-A3 are invalid in the intended interpretation, they are valid in other interpretations, so it is wrong to say that they are invalid *simpliciter*. The advantage of this reply is that it accounts for the apparent invalidity of A1-A3 in the intended interpretation. Its disadvantage, however, is that the relation between A1-A3 and S1-S3 becomes problematic. On the standard understanding of formal validity, an argument form is valid if and only if all its instances are valid. Assuming that validity is a property of interpreted arguments, this is to say that an argument form is valid if and only if all its instances are valid interpreted arguments. But then it turns out that some valid argument forms, S1-S3, have invalid instances, which is quite hard to accept.[18]

What has been said so far shows that it is not clear how the advocates of contextualism can reject the assumption that A1-A3 are invalid. Of course, rejecting that assumption is not the only way to deny the absurdity of the alleged consequence that A1-A3 instantiate S1-S3. The other way is to reject the assumption that S1-S3 are valid. However, such a reply throws the baby out with the bathwater. To say that S1-S3 are invalid is to deny the basic principles of modal logic. For the validity of S1-S3 follows from those principles. If S1-S3 are invalid, then the semantics of the language in which they are expressed is not the familiar semantics of modal logic, and $\Box$ does not have its familiar meaning. Even if one is willing to accept this consequence, which is not easy to swallow, the question remains of how the thesis that counterfactuals are strict conditionals can be maintained in some sense that matters to the Stalnaker-Lewis argument. For that argument is intended to establish that counterfactuals

[17] This line of argument is developed in different ways in Yagisawa 1993, Iacona 2010, and Georgi 2015.

[18] Note that the case of a valid form with invalid instances significantly differs from the case considered in section 2 of an invalid form with valid instances.

aren't strict conditionals just in the familiar sense.

Since the advocates of contextualism can hardly deny the absurdity of the alleged consequence of the thesis that counterfactuals are strict conditionals, it seems that a better option for them is to deny that the thesis has that consequence. As it turns out from section 2, this is the kind of reply provided by ellipticism. However, there are significant differences at the formal level. If counterfactulas are strict conditionals whose strictness varies as a function of their explicit antecedent, the obvious way to formally represent their variablity is to adopt indexed necessity operators $\Box_i$, where each i bears some relation to the antecedent of the formula in which it occurs. This way it can be contended that A1-A3 do not instantiate S1-S3 but invalid schemas in which different indices occur. Although there is nothing intrinsically wrong with this option, its formal part need be developed in order to be properly assessed, as it departs to some extent from standard modal logic. Ellipticism implies nothing like that, since S1-S3 are replaced by invalid schemas in the same language, S4-S5. So it seems that the best reply to the Stalnaker-Lewis argument that the advocates of contextualism can offer is a logically more complex variant of the ellipticist reply.

5 Disjunctive antecedents

This last section shows how ellipticism can handle the old problem of disjunctive antecedents. The problem concerns the inference schema called *simplification of disjunctive antecedents*, or SDA:

SDA If p or q had been the case, then r would have been the case.
$\therefore$ If p had been the case, then r would have been the case.

On the one hand, it may seem that SDA is a valid schema, for there are clear cases in which we reason in accordance with it. Consider the following sentence:

(17) If either Oswald had not fired or Kennedy had been in a bullet-proof car, Kennedy would be alive today.

What (17) conveys is that each of two possible events, Oswald not firing and Kennedy being in a bullet-proof car, would have lead to

the same result independently of the other, Kennedy being alive today. So it seems that from (17) we can infer

(18) If Oswald had not fired, Kennedy would be alive today.

And the same goes for the other disjunct.[19]

On the other hand, it has been argued that SDA is invalid, in that there are clear counterexamples to it. Suppose someone asks which side Spain fought on in World War II, and we reply that Spain did not enter the war, then adding the following sentence:

(19) If Spain had fought on the Axis side or on the Allied side, she would have fought on the Axis side.

In this case what we definitely are not willing to infer

(20) If Spain had fought on the Allies side, she would have fought on the Axis side.[20]

When uttering (19), we don't want to say that each of two possible events, Spain fighting on the Axis side and Spain fighting on the Allies side, would have lead to the same result independently of the other, Spain fighting on the Axis side. Rather, we want to say that if the disjunction 'Spain fought on the Axis side or Spain fought on the Allied side' were true, it would be true in virtue of the first disjunct. Therefore, not every counterfactual 'If *p* or *q* had been the case, then *r* would have been the case' is like (17).

According to the Stalnaker-Lewis view, SDA is invalid. If one represents the premise as $(\alpha \vee \beta) > \gamma$ and the conclusion as $\alpha > \gamma$, one gets an invalid argument form: it may be the case that $(\alpha \vee \beta) > \gamma$ is true, because every relevantly similar β-world is a γ-world, while $\alpha > \gamma$ is false. The friends of the Stalnaker-Lewis view have provided at least two arguments against the validity of SDA. The first goes as follows. Inferences such as that from (17) to (18) are indeed plausible. But their plausibility can be explained without assuming that SDA is a valid schema. Although it might seem that (17) has the form $(\alpha \vee \beta) > \gamma$, in reality it has the form $(\alpha > \gamma) \wedge (\beta > \gamma)$, hence (18) amounts to one of its conjuncts. The word 'or' in (17) is not to be

[19] Fine 1975:453, Nute 1975:775-776, Ellis, Jackson and Pargetter 1977:355.

[20] The example comes from McKay and Van Inwagen 1977.

read in the standard way, as it often happens. Sometimes the surface structure of natural language is misleading.[21]

This argument is not very convincing. It is legitimate to suppose that the plausibility of the inference from (17) to (18) can be explained without assuming that SDA is a valid schema. But the claim that (17) has the form $(\alpha > \gamma) \wedge (\beta > \gamma)$ requires an independent justification, and it is not clear that such a justification can be provided. The trouble is not only the weakness of the evidence for that claim, but also the strength of the evidence against it. As it has been observed, 'or' seems to behave in the usual way when negated. Consider the following sentence:

(21) If it had not been the case that either Oswald had not fired or Kennedy had been in a bullet-proof car, Kennedy would not be alive today.

Prima facie, (21) is equivalent to 'If it had been the case that Oswald had fired and Kennedy had not been in a bullet-proof car, Kennedy would not be alive today'. This is exactly what we should expect given the standard assumption that $\sim(\alpha \vee \beta)$ is equivalent to $\sim\alpha \wedge \sim\beta$. Instead, if the logical form of (17) were $(\alpha > \gamma) \wedge (\beta > \gamma)$, its logical form would be something like $\sim((\alpha > \sim\gamma) \wedge (\beta > \sim\gamma))$ or $\sim(\alpha > \sim\gamma) \wedge \sim(\beta > \sim\gamma)$, which is quite implausible.[22]

The second argument is that the assumption that SDA is a valid schema, combined with the apparently innocuous principle of substitution of equivalents, leads to undesirable results. Since α is equivalent to $(\alpha \wedge \beta) \vee (\alpha \wedge \sim\beta)$, by substitution of equivalents we get that $\alpha > \gamma$ entails $((\alpha \wedge \beta) \vee (\alpha \wedge \sim\beta)) > \gamma$. But if SDA is a valid schema, from $((\alpha \wedge \beta) \vee (\alpha \wedge \sim\beta)) > \gamma$ we get $(\alpha \wedge \beta) > \gamma$. So it turns out that $\alpha > \gamma$ entails $(\alpha \wedge \beta) > \gamma$, which is the unacceptable rule of strengthening the antecedent.[23]

This argument can have some effect only on those who accept the formalization suggested by the Stalnaker-Lewis view, hence reject the thesis that counterfactuals are strict conditionals. For if the

[21] Loewer 1976: 534-537, McKay and Van Inwagen 1977: 355, Lewis 1977: 360-361.

[22] See Ellis, Jackson and Pargetter 1977: 356.

[23] Fine 1975: 453, Lewis 1977: 359.

thesis holds, no such trouble can arise. According to a strict conditional analysis, the most natural formal counterpart of SDA is a valid argument form:

S7 $\Box((\alpha \vee \beta) \supset \gamma)$
$\therefore \quad \Box(\alpha \supset \gamma)$

Since every α-world is a $\alpha \vee \beta$-world, if every accessible $\alpha \vee \beta$-world is a γ-world, every accessible α-world must be a γ-world. Assuming substitution of equivalents, from S7 we get that $\Box(\alpha \supset \gamma)$ entails $\Box((\alpha \wedge \beta) \supset \gamma)$. But there is nothing wrong with that, since S1 is valid.

From the two arguments considered emerges no straightforward solution to the problem of disjunctive antecedents. It is reasonable to say that SDA is not a valid schema, in that not every sentence that may occur as a premise of SDA is like (17). Undoubtedly, a distinction must be drawn between counterfactuals such as (17) and counterfactuals such as (19). But it would be nice to have an explanation of this distinction that does not rest on highly debatable assumptions. Ellipticism can provide such explanation.

Consider (17). In this case it is said that if each of the disjuncts that occur in the antecedent were true, it would make the consequent true. Accordingly, (17) is properly phrased as follows: necessarily, if Oswald has fired and things are relevantly like in the actual world or Kennedy has been in a bullet-proof car and things are relevantly like in the actual world, then Kennedy is alive today. So its formal representation is $\Box((\alpha \vee \beta) \supset \gamma)$, where α stands for 'Oswald has fired and things are relevantly like in the actual world', and β stands for 'Kennedy has been in a bullet-proof car and things are relevantly like in the actual world'. Since the logical form of (18) is $\Box(\alpha \supset \gamma)$, (17) entails (18) in virtue of S7.

Now consider (19). In this case it is said that if the disjunction that forms the antecedent were true, the consequent would make it true. Accordingly, (19) is properly phrased as follows: necessarily, if Spain fought either on the Axis side or on the Allied side and things are relevantly like in the actual world, then Spain fought on the Axis side. So its formal representation is $\Box(\alpha \supset \beta)$, where α stands for 'Spain fought either on the Axis side or on the Allied side and things

are relevantly like in the actual world', and β stands for 'Spain fought on the Axis side'. Since the formal representation of (20) is $\Box(\gamma \supset \beta)$, where γ stands for 'Spain fought on the Allies side and things are relevantly like in the actual world', the inference from (19) to (20) is not formally valid.

More generally, SDA is ambiguous. There are cases of SDA in which the premise is adequately represented as a strict conditional with a disjunctive antecedent, and cases of SDA in which the premise does not have that form. The inferences of the first kind are valid because they instantiate S7. Those of the second kind are invalid because they instantiate an invalid argument form.

This explanation, just like that proposed by the friends of the Stalnaker-Lewis view, implies that there is no strict rule for the formalization of a sentence 'If *p* or *q* had been the case, then *r* would have been the case'. The recipe adopted so far for counterfactuals whose antecedent is a simple sentence or a conjunction works for (19) but not for (17). However, the account of (17) suggested entails no drastic revision of its apparent structure. This turns out clear if we consider the relation between (17) and (21). If (17) is represented as $\Box((\alpha \vee \beta) \supset \gamma)$, there is no trouble with the negation of its explicit antecedent. (21) can be represented as $\Box(\sim(\alpha \vee \beta) \supset \sim\gamma)$, which is equivalent to $\Box((\sim\alpha \wedge \sim\beta) \supset \sim\gamma)$ on the usual understanding of 'or'. This means that the logical form of (17) and the logical form of (21) turn out to be related exactly in the way one would expect.

Andrea Iacona
Center for Logic, Language, and Cognition
Department of Philosophy and Education
University of Turin
Via S. Ottavio 20, 10124 Turin
andrea.iacona@unito.it

References

Åqvist, L. 1973. Modal Logic with Subjunctive Conditionals and Dispositional Predicates. *Journal of Philosophical Logic* 2: 1-76.

Arlo-Costa, H. 2007. The Logic of Conditionals, *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2014/entries/logic-conditionals/>.

Brogaard, B. and Salerno, J. 2008. Counterfactuals and Context. *Analysis* 68: 39-46.
Burks, A. W. 1951. The Logic of Causal Propositions. *Mind* 60: 363-382.
Cross, C. B. 2011. Comparative World Similarity and What is Held Fixed in Counterfactulas. *Analysis* 71: 91-96.
Ellis, B.; Jackson, F. and Pargetter, R. 1977. An Objection to Possible-World Semantics for Counterfactual Logic. *Journal of Philosophical Logic* 6: 355-357.
Fine, K. 1975. Critical notice, 'Counterfactuals'. *Mind* 84: 451-458.
Georgi, G. 2015. Logic for Languages Containing Referentially Promiscuous Expressions. *Journal of Philosophical Logic* 44: 429-451.
Gillies, A. S. 2007. Counterfactual Scorekeeping. *Linguistics and Philosophy* 30: 329-360.
Iacona, A. 2010. Truth Preservation in any Context. *American Philosophical Quarterly* 47: 191-199.
Iacona, A. 2011. Counterfactual Fallacies. *Humana.Mente* 19 :1-9.
Kaplan, D. 1989. Demonstratives. In *Themes from Kaplan*, ed. by J. Perry J. Almog and H. Wettstein. Oxford University Press, 481-563.
Leibniz, G. W. 1985. *Theodicy*. Open Court.
Lewis, D. 1973. *Counterfactuals*. Blackwell.
Lewis, D. 1977. Possible World Semantics for Counterfactual Logics: A rejoinder. *Journal of Philosophical Logic* 6: 359-363.
Loewer, B. 1976. Counterfactuals with Disjunctive Antecedents. *Journal of Philosophy* 73: 531-537.
Lowe, E. J. 1990. Conditionals, Context, and Transitivity. *Analysis* 50: 80-87.
Lowe, E. J. 1995. The Truth about Counterfactuals. *Philosophical Quarterly* 45: 41-59.
McKay, T. and Van Inwagen, P. 1977. Counterfactuals with Disjunctive Antecedents. *Philosophical Studies* 31: 353-356.
Nute, D. 1975. Counterfactuals and the Similarity of Worlds. *Journal of Philosophy* 72: 773-778.
Sider, T. 2010. *Logic for Philosophy*. Oxford University Press.
Stalnaker, R. 1991. A Theory of Conditionals. In *Conditionals*, ed. by F. Jackson. Oxford University Press, 28-45.
Tichý, P. 1984. Two Parameters vs. Three. *Philosophical Studies* 45: 147-179.
von Fintel, K. 2001. Counterfactuals in a Dynamic Context. In *A Life in Language*, ed. by M. Kenstowicz and Ken Hale. MIT Press, 123-152.
Warmbrod, K. 1981. Counterfactuals and Substitution of Equivalent Antecedents. *Journal of Philosophical Logic* 10: 267-289.
Yagisawa, T. 1993. Logic Purified. *Nous* 27: 470-486.

Rightness = Right-Maker: Reduction or *Reductio*?

Joseph Long
SUNY, The College at Brockport

BIBLID [0873-626X (2015) 41; pp. 193-206]

Abstract
I have recently argued that if the causal theory of reference is true, then, on pain of absurdity, no normative ethical theory is true. In this journal, Michael Byron has objected to my *reductio* by appealing to Frank Jackson's moral reductionism. The present essay defends my *reductio* while also casting doubt upon Jackson's moral reductionism.

Keywords
Causal theory of reference, right-making properties, moral reductionism, Frank Jackson, justifying reasons.

In "Right-making and Reference", I argue that if the causal theory of reference is true, then, on pain of absurdity, no normative ethical theory is true (Long 2012). The causal theory of reference (CTR, henceforth) holds that a term 'T' rigidly designates a property *F* iff the use of 'T' by competent users of the term is causally regulated by *F*.[1] For example, since being H_2O causally regulates the competent use of 'is water', 'is water' rigidly designates being H_2O. A normative ethical theory, by contrast, is a theory that attempts to specify which property or properties are the fundamental right-making properties (FRM-properties, henceforth). A property is an FRM-property iff it is purely descriptive and is such that, if possessed by a right action, is what ultimately explains the action's being right. For example, utilitarianism implies that there is exactly one FRM-property, viz., maximizing aggregate pleasure: According to utilitarianism, maximizing aggregate pleasure is what makes all and only right actions right. Since a normative ethical theory attempts to specify

[1] See, e.g., Boyd 1988, Kripke 1980, and Putnam 1975.

Disputatio, Vol. VII, No. 41, November 2015
Received: 15/04/2015 Revised: 22/07/2015 Accepted: 08/08/2015

which purely descriptive properties are FRM-properties, then if no property is an FRM-property, no normative ethical theory is true. I argued that CTR implies, on pain of absurdity, that no property is an FRM-property and, thus, that no normative ethical theory is true. In this journal, Michael Byron (2014) has objected to my *reductio* by appealing to Frank Jackson's moral reductionism. The present essay defends my *reductio* while also casting doubt upon Jackson's moral reductionism.

1 A reductio

I begin with a summary of my earlier argument, which relies upon the following two assumptions:

(A1) A property is an FRM-property only if the moral property of being right exists.

(A2) The moral property of being right exists only if our term 'is right' refers to it.

Regarding the first assumption, if the property of being right does not exist, then no property can make an action right, in which case no property can be right-making. Thus, (A1). As for (A2), its denial is this:

(~A2) The moral property of being right exists, but our term 'is right' does not refer to it.

Claiming (~A2) amounts to denying that the relation between 'is right' and being right is a reference relation, which denial would undermine CTR's motivation. So, for the purposes of this essay, we can assume (A2). With (A1) and (A2) in hand, here is my argument in truncated form:

(P1) There is a true normative ethical theory only if there is an FRM-property.

(P2) If there is an FRM-property, then it causally regulates the competent use of 'is right'.

(P3) If an FRM-property causally regulates the competent use of

'is right', then, assuming (A1) and (A2), CTR implies that the FRM-property is identical to the property of being right.

(P4) An FRM-property's being identical to the property of being right entails absurdity.

∴ (C) Either no normative ethical theory is true, or CTR is false.

As construed, the argument is valid. So, let us consider each premise. We have already seen the argument for (P1): since a normative ethical theory attempts to specify which purely descriptive properties are FRM-properties, no such theory is true if there is no FRM-property.

Premise (P2) results from an inductive inference. Suppose, for ease, that there is exactly one FRM-property, in which case 'is right' applies to all and only actions possessing the FRM-property. If 'is right' applies to all and only actions possessing the FRM-property, then the competent use of 'is right' at least "tracks" the FRM-property. For example, if maximizing aggregate pleasure is the one and only FRM-property, then the competent use of 'is right' "tracks" maximizing aggregate pleasure. Presumably, the best explanation of this tracking behavior is that the FRM-property causally regulates the competent use of 'is right'. So, (P2) is probably true.

Turning to (P3), trivially an FRM-property causally regulates the competent use of 'is right' only if an FRM-property exists. According to (A1), an FRM-property exists only if the property of being right also exists. So, given (A1), it follows that if an FRM-property causally regulates the competent use of 'is right', then the property of being right exists. Now, according to (A2), if the property of being right exists, then our term 'is right' refers to it. So, together (A1) and (A2) imply that if an FRM-property causally regulates the competent use of 'is right', then 'is right' refers to the property of being right. But CTR implies that if an FRM-property causally regulates the competent use of 'is right', then 'is right' rigidly designates the FRM-property, which in turn implies that the FRM-property is identical to being right. Therefore, (P3): together (A1), (A2), and CTR imply that if an FRM-property causally regulates the competent use of 'is right', then the FRM-property is identical to the

property of being right.

According to (P4), however, an FRM-property's being identical to being right entails absurdity. My main support for (P4) is that,

(P4*) The "property that *explains* an action's being right cannot be *identical* to the property of being right" (2012: 278).

2 Jackson's moral reductionism

Byron, however, objects. The objection as I understand it has two parts: the first aims at casting doubt upon (P4*), while the second tries to show that (P4*) is actually false. To cast doubt upon (P4*), Byron essentially shows that the following universal statement, of which (P4*) is an instance, is false:

(UI) For any two properties *F* and *G,* the *F* that explains *x*'s having *G* cannot be identical to *G*.

Here is a counterexample to (UI): the property of being an *Apatosaurus* explains an organism's being a *Brontosaurus*, but being an *Apatosaurus* is identical to being a *Brontosaurus*.[2] Indeed, the property of being an *Apatosaurus* explains an organism's being a *Brontosaurus* precisely because being an *Apatosaurus* is identical to being a *Brontosaurus*. So, (UI) is false. But showing that (UI) is false shows only that, for *some* properties *F* and *G*, it is possible that $F = G$ and having *F* explains having *G*. It might be that the particular explanatory relation cited in (P4*) between the property that explains an action's being right and the property of being right prevents identifying these *particular* properties with each other. So, showing that (UI) is false might—if anything—make one suspicious of (P4*), but anything more than mere suspicion is unwarranted. Consequently, Byron needs to address (P4*) specifically.

In the second part of his objection, Byron tries to show that (P4*)

[2] Byron makes the same point in terms of being the morning star and being the evening star (Byron 2014: 142); however, putting the point in terms of being a *Brontosaurus* and being an *Apatosaurus* would better support Byron's point, since the present discussion is about property-identity rather than object-identity. (It is worth noting that whether the natural kinds *Brontosaurs* and *Apatosaurus* are identical has just come into question; see Tschopp et al. 2015.)

is false. To do so, Byron appeals to Frank Jackson's (1998) moral reductionism.[3] Here is how Byron describes Jackson's view. First, as Byron rightly states, Jackson's view holds that "normative properties are reducible to descriptive properties because the former constitute a proper subset of the latter" (Byron 2014: 142).[4] Furthermore, as Byron claims, "Jackson defines descriptive properties as those that can be picked out by descriptive predicates" (Byron 2014: 142). In conclusion, Byron quotes Mark Schroeder as saying, Jackson's reductionism "'amounts to the claim that normative properties can be picked out by uncontroversially descriptive predicates. *This* is a perfectly coherent view'" (Byron 2014: 142; Schroeder 2003: 10; emphasis in the original). What is more, claims Byron, Jackson's reductionism can "underwrite" the explanatory relation between being right and the FRM-property to which being right is identical (Byron 2014: 142-143). For, if—as Jackson's view implies—being right is a proper subset of purely descriptive properties, then should we discover that an FRM-property term picks out that proper subset, we can conclude that the FRM-property term's referent—that is, the FRM-property—is identical to being right.[5] "Far from being impos-

[3] Byron initially considers an objection according to which, basically, a property *F* could be both an FRM-property and identical to being right since (i) *F*'s being an FRM-property could amount to *F*'s playing the right-making role, (ii) *F*'s playing the right-making role could amount to *F*'s constituting the property of being right, and (iii) property-constitution could be a form of property-identity. I set aside this objection by Byron for two reasons. First, there are good reasons, none of which Byron addresses, to doubt that property-constitution could be a form of property-identity (see, e.g., Baker 2007: 111-116; Brink 1989: 157-158). But, second, given his appeal to Jackson's moral reductionism, which does not invoke property-constitution, Byron is able to avoid thorny questions about property-constitution altogether.

[4] Relevant to n. 3 above, the term 'constitutes' in the quote from Jackson does not refer to a relation between particular properties. Indeed, as far as I know, Jackson never invokes property-constitution to describe the relation between two particular properties.

[5] This is a charitable interpretation of Byron. Literally, Byron has us first suppose that "value-maximizing is the (descriptive) FRM, and that Jackson is right to think that the normative property of rightness is reducible to a descriptive property" (2014: 143). Byron then claims, "It follows that…rightness is [identical to] value-maximizing" (2014: 143). But just because rightness reduces to some

sible or absurd as Long claims, that result would be informative and illuminating" (Byron 2014: 143).

To evaluate Byron's argument, we must recognize, first, that Jackson's reductionism does not *merely* amount to "the claim that normative properties can be picked out by uncontroversially descriptive predicates," as Byron quotes Schroeder as asserting. For, if that were all that Jackson's reductionism amounted to, then Jackson's view would also imply that normative properties are reducible to geometrical-shape properties since one could use a geometrical-shape property-term—'is a triangle', for example—to pick out the normative property of, say, being right. But showing that one could use 'is a triangle' to refer to being right does not show that being right is reducible to being a triangle; it shows only that one can use 'is a triangle' equivocally. To avoid counting the equivocal use of a term as a form of reduction, Jackson's view needs to show that the property of being right could turn out to be identical to an FRM-property regardless of which terms refer to which properties.

As it turns out, Jackson's view of properties purports to do precisely this. On Jackson's view, properties are basically sets of possible objects.[6] For example, the property of being a triangle would be the set of all possible triangles; being a *Brontosaurus* would be the set of all possible *Brontosauruses*; and being right would be the set of all possible right actions. Now, presumably every possible right action possesses some purely descriptive property; however, some possible actions with a purely descriptive property are not right actions. Therefore, if properties are sets, then being right will turn out to be a proper subset of the union of purely descriptive properties. Of course, if properties are sets, then an FRM-property is itself a set: the set of all possible actions with the FRM-property. But all and only right actions have an FRM-property. So, should properties turn out to be sets, then any FRM-property will be a subset of being right: If there are multiple FRM-properties, then each FRM-property will be a proper subset of being right; and if there is exactly

purely descriptive property, it does not follow that rightness reduces specifically to the FRM-property.

[6] See Jackson 1998: 125-128. McNaughton and Rawling 2003 also contains a useful discussion Jackson's view of properties.

one FRM-property—perhaps maximizing expected hedonic value, to use Jackson's example—then the FRM-property will turn out to be identical to being right. Now, to be sure, Byron mentions that on Jackson's view being right is a proper subset of purely descriptive properties, and obviously being right could be such a subset only if being right is itself a set. But it needs to be emphasized that Jackson's view of properties *qua* sets is what allows Jackson to identify being right with an FRM-property. Consequently, here is how Byron's objection to (P4*) *should* go:

(1) It is coherent that,

(i) the property of being right is the set *R* of all and only possible right actions,

(ii) the property of maximizing expected hedonic value is the set *D* of all and only possible actions that maximize expected hedonic value, and

(iii) all and only members of *R* are also members of *D*.

(2) If (1), then being right could turn out to be identical to maximizing expected hedonic value.

(3) If being right could turn out to be identical to maximizing expected hedonic value, then an FRM-property can be identical to being right.

∴ (4) An FRM-property can be identical to being right.

If (4) is true, then the property that explains an action's being right can be identical to the property of being right, which is precisely what (P4*) denies. As construed, the argument is valid. Furthermore, I will grant premises (2) and (3) and argue against (1), to which I now turn.

3 Objecting to (1) and defending (P4*)

At first, one might be tempted to object to (1) on the grounds that it allows a property to be both normative and purely descriptive. For, if being right is identical to *R*, and maximizing expected hedonic value is identical to *D*, then if *R* and *D* are identical to each other, it will turn out that being right is normative iff maximizing expected hedonic value is normative and that maximizing expected hedonic value is purely descriptive iff being right is purely descriptive. But allowing a property to be both normative and purely descriptive, the objection would continue, obliterates the is/ought divide between properties.

Unfortunately for our would-be objector, this is not so much an objection as just a part of Jackson's view. For, as Byron rightly states, Jackson *defines* a normative property as a property that can be picked out by a normative property-term, and a purely descriptive property as a property that can be picked out by a purely descriptive property-term.[7] So, on Jackson's view, the is/ought divide is located at the level of property-*terms*. But if a normative property-term applies to all and only the members of a set of possible actions to all and only of which a purely descriptive property-term applies, then—again assuming that properties are sets—it follows that the set in question is a normative *and* purely descriptive property. By itself, that is no objection to Jackson's view; it is, rather, just part of the view, and that part is at least coherent.

Nonetheless, one might still challenge Jackson's view of normative properties *qua* sets of possible actions. There are two ways to do this: one can try to show that Jackson's view of properties is simply false, or, more modestly, one can argue that Jackson's view cannot adequately account for normative properties.[8] I will take the second tack. But I will also show that Jackson's view fails for the same

[7] See Byron 2014: 142 and Jackson 1998: 120-121.

[8] As an example of taking the first tack, see Elliott Sober 1982. Jackson considers a variation of Sober's case and responds (1998: 126-127). The second tack is more modest since Jackson's general view of properties could be true and yet fail to account for normative properties because normative properties do not exist (see, e.g., Mackie 1977 and Joyce 2006).

reasons that (P4*) is true. So, the argument I shall develop will simultaneously show that (1) is false and give support to my original (2012) argument.

As stated above, the particular explanatory relation between the property that explains an action's being right and the property of being right might make it impossible to identify these two particular properties with each other. I will now show why the explanatory relation between these two properties does indeed, as (P4*) claims, make it impossible to identify them with each other. First, consider that an action is right just in case it is justified. This is so presumably because being right and being justified, as properties of actions, are one and the same property—to be right just is to be justified. It is a platitude, furthermore, that actions are justified for reasons: if an action is justified, there is a reason it is justified. (Call such reasons 'justifying reasons'.[9]) Given that justifying reasons are what justify actions, we cannot identify a justifying reason with the fact that an action is justified. For, to do so would entail claiming this: that which justifies the action is identical to the fact that the action is justified. But that claim is incoherent. The fact that an action is justified cannot be that which justifies the action. It is worth noting that this sort of incoherence is not peculiar to justification. For example, it holds equally for explanation: That which explains an event cannot be identical to the fact that the event is explained. That an event is explained cannot be what explains the event. Similarly, that an action is justified cannot be what justifies the action. Since being right is identical to being justified, it thus follows that an action's justifying reason cannot be identical to the fact that the action is right. An action's justifying reason, that is, must be distinct from the fact that the action is right. Now, on what is probably the most common view of justifying reasons, a justifying reason is a fact that

[9] Justifying reasons should be distinguished from so-called explanatory reasons, the latter of which often appeal to the psychological states of the agent performing the action: the (explanatory) reason the agent performed that action is that (say) the agent had a certain belief-desire pair. The term 'explanatory reason' is infelicitous, given that justifying reasons can also figure into explanations—namely, they explain why an action is justified. Indeed, that justifying reasons are also explanatory in this way is important for the present argument.

counts conclusively in favor of an action.[10] For example, if the fact that an action maximizes expected hedonic value counts conclusively in favor of the action, then that fact is what justifies the action. But even if justifying reasons should be facts, a justifying reason cannot be identical to the fact that an action is right. For, whether or not a justifying reason is a fact, identifying an action's justifying reason with the fact that the action is right entails identifying that which justifies the action with the fact that the action is justified, which again is incoherent. The fact that an action is justified cannot be that which justifies the action. It follows, then, that whether or not justifying reasons are facts, we cannot, on pain of incoherence, identify an action's justifying reason with the fact that the action is right.

We are now in a position to see why, as per (P4*), we cannot identify the property that explains an action's being right with the property of being right and, thus, why (1) is false. Henceforth, let us assume the platitude that actions are justified for reasons—which, for ease, I shall take to be facts—and that being right is identical to being justified. From these two assumptions, we get our first premise:

(~1.1) An action is right only if a fact justifies the action.

Now, as explained above, it is incoherent to identify the fact that justifies an action with the fact that the action is right. So, here is our second premise:

(~1.2) If a fact justifies the action, then identifying the fact that justifies the action with the fact that the action is right is incoherent.

Our final premise is this:

[10] Theorists who either identify justifying reasons with facts or take facts to "give" justifying reasons include Broome (2004), Dancy (2000), Darwall (1983), McNaughton and Rawling (2003), Parfit (1997), Raz (1975), and Shafer-Landau (2003). For my purposes here, it will not make a difference whether justifying reasons are identical to facts or facts "give" justifying reasons. Also, I say 'conclusively' because reasons are often taken to be *pro tanto* whereas being justified implies success. If reasons are *pro tanto*, then a justifying reason is a consideration that counts in favor of an action and is not overridden by other considerations.

(~1.3) If identifying the fact that justifies the action with the fact that the action is right is incoherent, then identifying the property of being right with an FRM-property also leads to incoherence.

The first step toward seeing that (~1.3) is true requires seeing how FRM-properties relate to justifying reasons. Assuming (as we are) that justifying reasons are facts, we can express the relation like this: a property *F* is an FRM-property iff a token action's justifying reason is the fact in which the action possesses *F*. For example, if maximizing expected hedonic value is the one and only FRM-property, then what would justify an action would be the fact that the action maximizes expected hedonic value.

The second step toward seeing that (~1.3) is true requires recognizing that the following conditional is also true: If *F* is an FRM-property just in case a token action's justifying reason is the fact in which the action possesses *F*, then, on pain of incoherence, should *F* be an FRM-property, *F* cannot be identical to being right. To see why this conditional is true, suppose that a token action *a* possesses an FRM-property. If *a* possesses an FRM-property, then there is a fact in which *a* possess an FRM-property and, what is more, that fact justifies *a*. If the fact in which *a* possesses an FRM-property is what justifies *a*, then, trivially, some fact justifies *a*; and if some fact justifies *a*, then *a* is justified. So, *a*'s possessing an FRM-property results in there being two facts: the fact in which *a* possesses the FRM-property and the fact that *a* is justified. As explained above, however, we cannot identify the two facts. For, to do so would amount to claiming that that which justifies *a* is identical to the fact that *a* is justified, which is incoherent. So, the fact in which *a* possesses an FRM-property cannot be identical to the fact that *a* is justified. But the token action in both facts is one and the same action, viz., *a*. Consequently, if the properties in the two facts should also be one and the same property, then the facts themselves will be one and the same fact. To see this, suppose that it is a fact that a token organism *o* is a *Brontosaurus* and it is also a fact that *o* is an *Apatosaurus*. If being a *Brontosaurus* is identical to being an *Apatosaurus*, then the fact that *o* is a *Brontosaurus* and the fact that *o* is an *Apatosaurus* are the same fact—the fact just involves a property (viz., being a member of a

certain natural kind) picked out by two property-terms: 'is a *Brontosaurus*' and 'is an *Apatosaurus*'. But the facts are identical nonetheless. By parity of reasoning, then, if an FRM-property is identical to being justified, then the fact that justifies our token action *a* is identical to the fact that *a* is justified, which is incoherent. So, on pain of incoherence, no FRM-property can be identical to the property of being justified; and since being right and being justified are one and the same property, it follows that no FRM-property can be identical to being right. Premise (~1.3) follows. Having already established (~1.1) and (~1.2), we can now validly infer that identifying the property of being right with an FRM-property leads to incoherence. Since (1) implies, to the contrary, that identifying the property of being right with an FRM-property is coherent, we can conclude that (1) is false. Since (1) is false, Byron's objection to (P4*) fails; and since Jackson's view of properties, when applied to moral properties, implies (1), we can also conclude that Jackson's view of properties fails to account for moral properties, which ultimately casts a dubious light upon Jackson's moral reductionism.

Finally, we can see why (P4*) is true. The particular explanatory relation cited in (P4*) prevents identifying the property that explains an action *a*'s being right with the property of being right, since (i) the property that explains *a*'s being right is, roughly put, the property whose possession by *a* is what justifies *a*[11] and (ii) being right and being justified are one and the same property. For, if (i) the property that explains *a*'s being right is (roughly put) the property whose possession by *a* is what justifies *a* and (ii) being right and being justified are one and the same property, then to identify the property that explains *a*'s being right with the property of being right entails identifying that which justifies *a* with *a*'s being justified, which is incoherent. So, the particular explanatory relation, cited by (P4*), between the property that makes an action right and the property of being right makes it impossible to identify these two properties, in which case not only is (1) is false, but (P4*) is true.

[11] If justifying reasons are facts, it would be more precise to say this: the property that explains an action *a*'s being right is the property whose possession by *a* results in the fact that justifies *a*. This degree of precision is not required for the point being made in the text.

Conclusion

In an earlier article, I argued that if the causal theory of reference is true, then, on pain of absurdity, no normative ethical theory is true (Long 2012). Michael Byron has objected to my argument—specifically, to the premise I have labelled '(P4*)'—by appealing to Frank Jackson's moral reductionism. My defense of (P4*) is essentially that Byron fails to appreciate the particular explanatory relation, cited in (P4*), between the property that explains an action's being right and the property of being right and that by getting clearer on this relation, we see not only that Byron's objection fails, but that (P4*) is both true and calls into question Jackson's account of moral properties and thus Jackson's moral reductionism. What is more, we can once again conclude that if the causal theory of reference is true, then no normative ethical theory is true.[12]

Joseph Long
Department of Philosophy
SUNY, The College at Brockport
350 New Campus Dr.
Brockport, NY 14420
jlong@brockport.edu

References

Baker, Lynne Rudder. 2007. *The Metaphysics of Everyday Life: An Essay in Practical Realism.* Cambridge, UK: Cambridge University Press.

Boyd, Richard. 1988. How to Be a Moral Realist. In *Essays on Moral Realism,* edited by Geoffrey Sayre-McCord. Ithaca: Cornell University Press.

Brink, David. 1989. *Moral Realism and the Foundations of Ethics.* Cambridge, UK: Cambridge University Press.

Broome, John. 2004. Reasons. In *Reason and Value: Themes in the Moral Philosophy of Joseph Raz,* edited by R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith. Oxford: Oxford University Press.

Byron, Michael. 2014. Right-making, Reference, and Reduction. *Disputatio* 39: 139-45.

Dancy, Jonathan. 2000. *Practical Reality.* Oxford: Oxford University Press.

[12] I would like to thank Michael Byron for responding to my earlier article and Piers Rawling for discussing with me various and sundry topics relevant to this essay.

Dancy, Jonathan. 2004. *Ethics without Principles*. Oxford: Oxford University Press.
Darwall, Stephen. 1983. *Impartial Reason.* Cornell: Cornell University Press.
Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.
Joyce, Richard. 2006. *The Evolution of Morality*. Cambridge, MA: The MIT Press.
Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
Long, Joseph. 2012. Right-making and Reference. *American Philosophical Quarterly* 49: 277-80.
Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin.
McNaughton, David and Rawling, Piers. 2003. Descriptivism, Normativity, and the Metaphysics of Reasons. *Proceedings of the Aristotelian Society*, Supplementary Volume: 24-45.
Parfit, Derek. 1997. Reasons and Motivation. *Proceedings of the Aristotelian Society, Supplementary Volume:* 99-130.
Putnam, Hilary. 1975. The Meaning of Meaning. In *Mind, Language, and Reality: Philosophical Papers*, vol. 2. Cambridge, UK: Cambridge University Press.
Raz, Joseph. 1975. *Practical Reason and Norms.* Oxford: Oxford University Press.
Shafer-Landau, Russ. 2003. *Moral Realism: A Defense.* Oxford: Oxford University Press.
Sober, Elliott. 1982. Why Logically Equivalent Predicates May Pick out Different Properties. *American Philosophical Quarterly* 19: 183-9.
Tschopp, Emanuel *et al*. 2015. A Specimen-level Phylogenetic Analysis and Taxonomic Revision of *Diplodocidae* (*Dinosauria, Sauropoda*). *PeerJ*, https://peerj.com/articles/857/

Dummett's Legacy: Semantics, Metaphysics and Linguistic Competence

Massimiliano Vignolo
University of Genoa

BIBLID [0873-626X (2015) 41; pp. 207-229]

Abstract
Throughout his philosophical career, Michael Dummett held firmly two theses: (I) the theory of meaning has a central position in philosophy and all other forms of philosophical inquiry rest upon semantic analysis, in particular semantic issues replace traditional metaphysical issues; (II) the theory of meaning is a theory of understanding. I will defend neither of them. However, I will argue that there is an important lesson we can learn by reflecting on the link between linguistic competence and semantics, which I take to be an important part of Dummett's legacy in philosophy of language. I discuss this point in relation to Cappelen and Lepore's criticism of Incompleteness Arguments.

Keywords
Sematics, pragmatics, metaphysics.

1 Dummett's legacy: semantics and metaphysics

Throughout his philosophical career, Michael Dummett never gave up two main theses:

(I) The theory of meaning has a central and foundational place in philosophy.

(II) The theory of meaning is a theory of understanding.

Thesis (I) is the climax of the linguistic turn started with Frege and adopted later by logical positivists. It is the view that metaphysical issues must be resolved, or dissolved, by recourse to the theory of meaning. Contrary to positivists, who dismissed metaphysical issues either as nonsense or as issues concerning no matters of fact and reducible to pragmatic choices between different languages, Dummett

Disputatio, Vol. VII, No. 41, November 2015
Received: 07/01/2015 Revised: 27/07/2015 Accepted: 12/10/2015

reinterpreted metaphysical disputes as disputes concerning the truth conditions of sentences.[1] Whether one is justified to be a realist in some area of discourse depends on whether one is justified to assign realist truth conditions to sentences in that area of discourse, i.e. bivalent, epistemically transcendent truth conditions. Linguistic categories are also the starting point for the analysis of formal ontological notions. For example, the formal notion of *object* is to be understood in terms of the notion of reference of singular terms—with the notion of singular term to be explained on the basis of characteristic behaviour in syntactic and logical operations on sentences containing singular terms.[2] Dummett gave the philosophy of language a foundational role. If metaphysical issues are reformulated as questions about the structure and content of language, only the philosophy of language can provide the analysis of such structure.

Nowadays many, perhaps most, philosophers reject the foundational role of the philosophy of language and claim a substantive and autonomous role for metaphysics. They regard metaphysics as that part of the philosophical inquiry that is engaged to discover objective characteristics of reality and not the fundamental features of our thought about reality.

Thesis (II) is also central in Dummett's philosophy and struggle against semantic realism. According to Dummett, the theory of meaning must be tripartite in (a) a theory of reference, (b) a theory of sense and (c) a theory of force.[3] The theory of reference determines recursively the conditions for the application to each sentence of that notion which is understood as the central notion in the explanation

[1] See Dummett 1978: xl: 'The whole point of my approach to these problems [the disputes concerning realism] has been to show that the theory of meaning underlies metaphysics. If I have made any worthwhile contribution to philosophy, I think it must lie in having raised this issue in these terms.'

[2] See Dummett 1981. For a discussion of this point, see Wright 1983:53-64.

[3] See Dummett 1976: 127: 'Any theory of meaning was early seen as falling into three parts: the first, the core theory, the theory of reference; secondly, its shell, the theory of sense; and thirdly, the supplementary part of the theory of meaning, the theory of force... The theory of reference determines recursively the application to each sentence of that notion which is taken as central in the given theory of meaning... The theory of sense specifies what is involved in attributing to a speaker a knowledge of the theory of reference.'

of meaning. The theory of sense specifies what is involved in ascribing the knowledge of the theory of reference to speakers. The theory of sense is a theory of understanding that specifies that in which the knowledge of the theory of reference consists.[4] As the knowledge of the theory of reference is an implicit form of knowledge, the theory of sense must correlate the knowledge of each theorem of the theory of reference with a practical linguistic ability.[5] Dummett's criticism of semantic realism is that the classical notion of truth cannot serve as the central notion in the explanation of meaning, since it makes it impossible to construct a proper theory of sense. This is to say that one cannot specify what is involved in ascribing to speakers the implicit knowledge of the theorems of a classical two-valued semantics, which assigns epistemically transcendent truth conditions to sentences.

Dummett's argument against semantic realism is known as *The Manifestation Argument* and has the form of a *reductio*:[6]

1. Knowledge of meaning is knowledge of classical truth conditions.

2. Knowledge of meaning consists in the capacity to recognize, if appropriately placed, whether or not truth conditions obtain.

3. Classical truth conditions are such that, if actualized, they need not be recognizably so.

4. Knowledge of meaning is not knowledge of classical truth conditions.

[4] See Dummett 1975: 99: 'A theory of meaning is a theory of understanding.'

[5] See Dummett 1976: 72: 'We may therefore require that the implicit knowledge which he [the speaker] has of the theorems of the theory of meaning which relate to whole sentences be explained in terms of his ability to employ those sentences in particular ways... The ascription to him of a grasp of the axioms governing the words is a means of representing his derivation of the meaning of each sentence from the meanings of its component words, but his knowledge of the axioms need not be manifested in anything but the employment of the sentence.'

[6] I borrow this presentation of the manifestation argument from Tennant (1987).

According to premise 2, every speaker who knows the meaning of a sentence S must be able to recognize that its classical truth conditions obtain whenever they obtain. But S's classical truth conditions might obtain without being it possible to know that this is so. Therefore, there is no guarantee that the knowledge of S's meaning consists in a capacity that can be ever exercised. This is an absurd consequence, since to have a capacity is to be able to do something that can be done. Nobody possesses a capacity to do anything that cannot be done. Dummett drew the conclusion 4, which is the negation of premise 1, i.e. of semantic realism.

The argument rests on premise 2, which is a consequence of thesis (II). Thesis (II) is known as the *manifestation constraint* and is Dummett's explication of Wittgenstein's slogan that *meaning is use*. It expresses the view that the theory of meaning must include the theory of sense, which specifies that in which the knowledge of meaning consists.[7] Dummett said that a theory that meets the manifestation constraint specifies not only what speakers know, when they know the meanings of the expressions of the language they speak, but also that in which such knowledge consists, in such a way that one would acquire the knowledge of the meanings of the expressions of the language under study, were one taught the practical abilities that the theory of sense is called to describe.

The manifestation constraint has a constitutive import. It regards linguistic behaviour as something in need of analysis. Linguistic behaviour is analysed in order to determine the complex of linguistic abilities that constitute the mastery of the language. To know that a certain expression has a certain meaning is to be able to make a certain use of that expression and the theory of meaning must describe such patterns of use.

Some philosophers have rejected Dummett's Manifestation Argument by rejecting thesis (II), with its constitutive significance. They hold that the ascription of the implicit knowledge of the theory, which for each sentence specifies its classical truth-conditions, amounts to the ascription

[7] See, for example, Dummett 1977: 376: 'An argument of this kind is based upon a fundamental principle, which may be stated briefly, in Wittgensteinian terms, as the principle that a grasp of the meaning of an expression must be exhaustively manifested by the use of that expression and hence must constitute implicit knowledge of its contribution to determining the condition for the truth of a sentence in which it occurs; and an ascription of implicit knowledge must always be explainable in terms of what counts as a manifestation of that knowledge, namely the possession of some practical capacity.'

of internal states and allows for testable predictions about speakers' linguistic behaviour. They reject Dummett's manifestation constraint that semantics (theory of reference)—the core part of the theory of meaning—must be associated with a theory of understanding—the theory of sense—that provides an analysis of linguistic behaviour that isolates the patterns of linguistic abilities that *constitute* the implicit knowledge of the semantic theory.[8]

I will not defend Dummett's theses (I) and (II). I agree that there is a division in the philosophical labour between metaphysicians and philosophers of language, and that the philosophy of language does not have a foundational role in respect of other philosophical fields. I also agree that the Manifestation Argument can be blocked by rejecting the constitutive constraint. However, I will argue that there is a constraint that makes the link between linguistic competence and semantics more intimate than some philosophers believe. I take this constraint to be part of Dummett's legacy in the philosophy of language. I will address the point by discussing Cappelen and Lepore's criticism of Incompleteness Arguments. I will claim that despite the fact that they recognize a division in the philosophical labour between metaphysicians and philosophers of language, their criticism of Incompleteness Arguments is mistakenly grounded on an underestimation of the connection between linguistic competence and semantics.[9]

[8] Dummett goes on to argue that classical semantics is not adequate because there are no linguistic abilities that constitute implicit knowledge of epistemically transcendent truth conditions. See Dummett 1991: 303: 'A semantic theory may be criticised on the ground that it cannot be extended to a coherent or workable meaning-theory at all; and since, by definition, a semantic theory can be so extended, this criticism amounts to saying that it is not, after all, a genuine semantic theory.'

[9] It is worth noticing that I will not draw any conclusion against classical bivalent semantics. To the extent that I defend the Incompleteness Arguments against Cappelen and Lepore's criticism, I draw a conclusion against Minimalism in semantics, and in favor of Contextualism. I mention Dummett's view to argue that theoretical reflections on speakers' linguistic competence and linguistic practice put some constraints on semantics and that Minimalism does not satisfy such constraints. In this paper I use 'Minimalism' in the same sense as Cappelen and Lepore (2005: 1) use it. On Cappelen and Lepore's view there are few expressions that are context sensitive, and such expressions belong to the *Basic Set* of genuinely context sensitive expressions: indexicals ('I'), demonstratives ('that'),

2 Incompleteness arguments

Contextualists employ Incompleteness Arguments to maintain that certain expressions are context sensitive. Consider the following sentence:

(1) Bradley is tall.

An Incompleteness Argument starts from the premise that if one takes (1) in isolation from the information available in the context of utterance, then one is unable to truth evaluate (1). It is only if one takes account of contextual information that utterances of (1) are truth evaluable. For example, in the course of a conversation about the physical characteristics of presidential candidates, the utterance of (1) is true if and only if Bradley is 180 cm tall or over, i.e. tall in respect of the average height of the presidential candidates. Whereas in the course of a conversation about great NBA centers, the utterance of (1) is true if and only if Bradley is 205 cm tall or over, i.e. tall in respect of the average height of great NBA centers. This line of reasoning leads to the conclusion that there is no invariant proposition, i.e. the proposition that Bradley is tall *simpliciter*, which utterances of (1) express in all contexts. On the other hand, one has the intuition that there are both the proposition that Bradley is tall as compared with the class of the candidates to the presidency and the proposition that Bradley is tall as compared with the class of great NBA centers, which are the propositions expressed by utterances of (1) with the help of the information available in the context of utterance. In general, then, a successful Incompleteness Argument gives evidence that there is no invariant proposition that a sentence S expresses in all contexts of utterance. If, in addition, this conclusion

---

adverbs ('here'), adjectives ('actual') and contextuals ('enemy'). All semantic context sensitivity is grammatically (i.e. syntactically or morphemically) triggered. I use the term 'Contextualism' in a very broad sense which comprehends indexicalism à la Stanley (2007), according to which the Basic Set of genuinely context sensitive expressions is much larger than Cappelen and Lepore think, but all context sensitivity is linguistically triggered (in the logical form if not in the grammatical form) and pragmatism à la Carston (2002), Travis (2008), Recanati (2011), according to which the Basic Set is even larger and not all context sensitivity is linguistically triggered, but a large part of it involves free pragmatic processes.

is accompanied with the intuition that in each context of utterance S expresses a truth evaluable content relative to the contextual information, then an inference to the best explanation of that intuition leads to the conclusion that S (some expressions occurring in it) is context sensitive. For example, the intuition that the truth conditions of (1) and the propositions expressed by it vary, when the contexts of utterance vary, is explained within a theory that treats 'tall' as a context sensitive expression.

3 Cappelen and Lepore's criticism of incompleteness arguments

Cappelen and Lepore (2005) reject Incompleteness Arguments because, in their view, arguments of that kind aim at establishing a metaphysical conclusion about the existence of entities that might figure as constituents of propositions, like the property of being tall *simpliciter*, on the basis of psychological data. Psychological data, however, have no bearing on metaphysical issues. Cappelen and Lepore say that typically an incompleteness argument amounts to the following claim:

> Consider the alleged proposition that P that some sentence S semantically expresses. Intuitively, the world can't just be P *simpliciter*. The world is neither P nor not P. There's no such thing as P's being the case *simpliciter*. And so, there is no such proposition.
>
> So, for example, consider 'Al is ready'. Some authors contend that it is just plain obvious that there isn't any such thing as Al's being ready *simpliciter*. (Cappelen and Lepore 2005: 11)

Their presentation of incompleteness claims has unequivocally a metaphysical import. Cappelen and Lepore argue that those philosophers, who make use of Incompleteness Arguments to support Contextualism, are guilty of conflating metaphysical issues with linguistic ones. The data about speakers' dispositions to truth evaluate sentences in their contexts of utterance might be revelatory about psychological facts and facts about communication, but have no weight for metaphysical inquiries on what entities exist.

I claim that Cappelen and Lepore's criticism of Incompleteness Arguments reveals their misunderstanding of the real nature of such

arguments and, consequently, their underestimation of the real force of the arguments of that kind. Consider Taylor's illustration of an incompleteness argument. Discussing the structure of the semantic content of utterances of (2):

(2) It is raining.

Taylor says:[10]

> [(2)] is missing no syntactic sentential constituent, nonetheless, it is semantically incomplete. The semantic incompleteness is manifest to us as a felt inability to evaluate the truth value of an utterance of [(2)] in the absence of a contextually provided location (or range of locations). This felt need for a contextually provided location has its source, I claim, in our tacit cognition of the syntactically unexpressed argument place of the verb 'to rain'. (Taylor 2001: 61)

Leaving aside Taylor's own view about the semantics of the verb 'to rain', which goes along the lines of the *Hidden Indexical Theory*, Taylor's idea of incompleteness is that if a sentence gives rise to a *felt inability* to truth evaluate its utterances independently of contextual information, then the sentence contains some context sensitive expressions. As said above, Cappelen and Lepore's criticism is that an argument such as Taylor's must be rejected because psychological facts about how speakers feel about the truth evaluation of sentences have no weight on metaphysical questions about what entities exist.

4 The real goal of incompleteness arguments

I will not raise questions about the truth of Cappelen and Lepore's claim that psychological facts have no bearing on metaphysical questions. I will argue, instead, that the truth of this claim is beside the point, because an incompleteness claim is not a metaphysical claim on the existence of this or that entity. Incompleteness Arguments do

[10] The quotation from Taylor serves to highlight the idea that an incompleteness argument starts from a premise that registers the speakers' *felt inability* to truth evaluate some utterances independently of contextual information. Nothing in the quotation from Taylor gives evidence in favour of Cappelen and Lepore's reading of Incompleteness Arguments according to which a metaphysical conclusion about the existence of certain entities follows from that premise.

not provide evidence against the existence of certain entities, which might figure as constituents of propositions, but against the idea that such entities, if any, can be semantically associated with words as their semantic contents. I hold that an incompleteness claim is a significant claim in respect of linguistic competence and theoretical considerations about linguistic competence do have consequences for semantics (so I will argue). For example, the conclusion of an incompleteness argument concerning the adjective 'tall'[11] is not that the property of being tall *simpliciter* does not exist, because speakers are unable to truth evaluate the sentence (1) independently of contextual information. One might agree with Cappelen and Lepore that the existence and possibly the account of the property of being tall *simpliciter* is a matter for metaphysicians not for philosophers of language. I claim that the conclusion of the incompleteness argument is that a semantic theory, which assigned the property of being tall *simpliciter* to the adjective 'tall' as its semantic value, would be incompatible with any account of linguistic competence, according to which to learn the meaning of an expression and to be competent about its use is to be able to use that expression insofar as that expression is governed by a semantic norm (or by a semantic property with a normative import). Such a semantic theory could hardly have any theoretical interest for an overall theory of language use and linguistic behaviour. I shall elaborate on this point.

Cappelen and Lepore argue that the felt inability to truth evaluate a simple sentence like 'Bradley is tall' offers no positive evidence against the view that the property of being tall *simpliciter* exists and

[11] It is not the aim of this paper to defend contextualism about this or that expression. If one says to have the intuition that the sentences 'Bradley is tall' and 'the leaves are green' have determinate truth conditions independently of contextual information, that is fine to me with regard to the purpose of this paper and I will not argue to the contrary. The aim of this paper is to defend incompleteness arguments from Cappelen and Lepore's criticism. One might change the examples I discuss with others involving different sentences. Notice that Cappelen and Lepore do not question the premise that speakers are not able to evaluate certain sentences independently of contextual information. Thus, the reader is free to choose one of those sentences. Cappelen and Lepore grant that premise but argue that incompleteness arguments are illegitimate because they conflate premises that register psychological data with metaphysical conclusions. I argue that the conclusions of incompleteness arguments are not metaphysical at all.

is the semantic content of the adjective 'tall'. On the one hand, Cappelen and Lepore acknowledge that the question of giving an analysis of the property of being tall *simpliciter* or an account of what makes something tall *simpliciter* is a difficult problem, but one for metaphysicians, not for semanticists. On the other hand, Cappelen and Lepore (2005: 164) hold that semanticists have no difficulty at all to say which proposition the simple sentence 'Bradley is tall' expresses: it is the proposition *that Bradley is tall*. Nor have semanticists any difficulty to tell the truth conditions of the simple sentence 'Bradley is tall': *'Bradley is tall' is true if and only if Bradley is tall*.

I claim that Cappelen and Lepore's confidence in disquotational truth conditions betrays their underestimation of Incompleteness Arguments. A semantic theory for a language L aims to capture the semantic properties of the expressions of L. The point, which is relevant to our discussion, is that a semantic theory must be related to linguistic competence. This is so not only for those philosophers who hold that a semantic theory is a theoretical representation of the implicit knowledge of the language, which competent speakers possess. It is so also for those philosophers who reject the view that a semantic theory is a theoretical representation of what competent speakers implicitly know.[12] Indeed, a semantic theory for L cannot be fully assessed in isolation from questions related to how L-expressions are bestowed with their semantic properties and to what L-speakers typically do, whenever they are regarded as competent in the use of L, especially questions as to whether the linguistic abilities they manifest count as governed by semantic normative principles.

[12] See, for example, Devitt 1981: 93: 'What need explaining, basically, are the verbal parts of human behaviour. In explaining these, we must attribute certain properties (for example, being true and referring to Socrates) to the sounds and inscriptions produced, and certain other properties (for example, understanding "Socrates") to the people who produce those sounds and inscriptions.' See also Devitt 1999: 169: 'Linguistic competence is a mental state of a person, posited to explain his linguistic behaviour; it plays a key role—although not, of course, the only role—in the production of that behaviour. Linguistic symbols are the result of that behaviour; they are the products of the competence, its outputs… A theory of a part of the production of linguistic symbols is not a theory of the products, the symbols themselves. Of course, given the causal relation between competence and symbols we can expect a theory of the one to bear on a theory of the other. But that does not make the two theories identical.'

Suppose a semantic theory for, say, English contains a disquotational principle like the following, which arguably captures what Cappelen and Lepore have in mind, when they say that the semantic content of 'tall' is the property of being tall *simpliciter* and that semanticists have not difficulty at all to tell the truth conditions of 'Bradley is tall' and which proposition it expresses:

> (A) For any object *o*, 'tall' applies in English to *o* if and only if *o* is tall.

The point I want to stress is that it is theoretically significant for that semantic theory that an account is available about how the linguistic abilities of competent speakers count as governed by the principle (A). It is also theoretically significant that an account is available about how it comes that the word 'tall' has the semantic property of applying to all and only tall *simpliciter* objects. If there is evidence that no account of that kind is available, then there is evidence that the semantic theory in question is on a wrong track. As Michael Devitt (2007: 52) says, semantic contents are not "God given", but as conventions need to be established and sustained by regular uses. Words cannot have the semantic contents they have independently of the linguistic behaviour of competent speakers. Otherwise, it is impossible to explain how words get associated with their semantic properties and how such associations are learned (and transmitted) by being exposed to the linguistic practice. Moreover, a semantic theory that does not enable us to describe the linguistic behaviour as subject to semantic principles with a normative import is scarcely of any interest for an overall account of language use.

I claim that the gist of Incompleteness Arguments is not that certain entities, such as the property of being tall *simpliciter*, do not exist. Rather, it is that such entities, if any, cannot be the semantic contents of words. A semantic theory that assigned such entities to words, as their semantic contents, would be incompatible with any plausible account of language learning and language understanding, according to which by learning and understanding a language, we learn and understand expressions as governed by semantic principles with normative import.

Consider one of Travis' (1997) favourite examples. A speaker utters the sentence (3):

(3) The leaves are green.

speaking of a Japanese maple, whose leaves are naturally russet but have been repainted green. In a context of utterance in which the speaker talks with a photographer, who looks for a green subject, the speaker is taken to tell the truth. In another context of utterance in which the speaker talks with a botanist, who is interested in the natural colour of the plant, the speaker is not taken to tell the truth. The point that an incompleteness argument brings out is that competent speakers feel unable to truth evaluate utterances of the sentence (3) independently of the information available in the context of utterance. This result means that the linguistic abilities that are required for the mastery of the word 'green' cannot be construed as governed by the semantic norm expressed by the following disquotational principle:

(B) For any object *o*, 'green' applies in English to *o* if and only if *o* is green.

The reason why linguistic competence cannot be so construed is that the linguistic practice cannot be guided by such principle. As a matter of fact, the principle (B) states conditions of the application for 'green' that competent speakers are never able to track, as testified by their felt inability to truth evaluate sentences such as (3) independently of contextual information. To put it another way, the principle (B) specifies the semantic content of the word 'green'. Hence, the principle (B) states a norm about the use of 'green': it is correct to apply 'green' to all and only green *simpliciter* objects. Incompleteness Arguments show that the norm that the principle (B) states is not applicable, because nobody in the linguistic community is able to tell when it applies and when it does not. Since norms must be applicable, the conclusion follows that the principle (B) states no norms at all and, therefore, cannot be a semantic principle. The principle (B) does not play the normative role that is constitutive of semantic principles.

The consequence of Cappelen and Lepore's view is more radical and damaging than the view held by externalists such as Putnam (1975). Externalists hold that semantic properties are objective in the sense that words have their semantic properties independently

of explicit knowledge and discriminating abilities, which speakers or the linguistic community as a whole possess. In 1750, 'water' in Twin Earthian English referred to XYZ even though nobody knew the chemical composition of the liquid stuff on Twin Earth and nobody could discriminate XYZ from H_2O. Externalism has the consequence that semantic norms might elude even the most expert speakers of the community. In 1750, nobody could have been in a position to correct an application of the Twin Earthian word 'water' to H_2O. Had a Twin Earthian speaker talked to an Earthian speaker, they would have misunderstood each other, one speaking of XYZ and the other of H_2O. As Marconi (1997: 88) remarks, that would be a misunderstanding of a very peculiar kind, since nobody in the linguistic community could have pointed it out.

It is not my interest here to take side with externalists and defend their view from Marconi's criticism. Rather, my interest is to highlight the difference between externalism and the radical position that issues from Cappelen and Lepore's view. Externalists hold that semantic properties are unaffected by explicit knowledge and discriminating abilities. Semantic properties are determined by certain factual, causal connections to the world. Externalists, however, do have an account of how words are bestowed with their semantic properties, which rests on baptismal ceremonies and, above all, multiple groundings. A word has the reference it has because most significant referential practices, as a matter of fact, are related to that reference. This means that there are favourable—contextually favourable, not epistemically or cognitively favourable—circumstances in which Twin Earthian competent speakers believe, and believe it truly, that the conditions for the application of 'water' are satisfied. This confers the following principle:

(C) 'water' refers in Twin Earthian English to XYZ.

its normative role, although it might elude even the most expert speakers in the whole community, when they are not in a contextual favourable position (say an expert Twin Earthian speaker has been transported to Earth). Therefore, there are favourable circumstances in which Twin Earthian competent speakers are disposed to truly assent to the sentence 'that is water' and to correctly truth evaluate other sentences containing the word 'water'.

Incompleteness Arguments show that competent speakers are never disposed to truth evaluate sentences containing certain words independently of contextual information. For example, there are no circumstances in which competent speakers are disposed to truth evaluate 'Bradley is tall' independently of contextual information. This means that competent speakers are never able to track instances of the property of being tall *simpliciter*. This fact prevents any semantic theory from ascribing the property of being tall *simpliciter* to the adjective 'tall' as its semantic content by means of the principle (A), because competent speakers are never able to tell when the conditions for the application of 'tall', as captured by the principle (A), are satisfied. Such semantics is not compatible with any account of how the adjective 'tall' is bestowed with its semantic property and of how such semantic property exerts a normative role over the linguistic practice.

5 Cappelen and Lepore's charge of verificationism

Cappelen and Lepore (2005: 164-5) take into consideration this form of resistance to their rejection of Incompleteness Arguments. They respond that semantics is not in the business of telling what the world is like. Therefore, semantics is not in the business of telling whether, say, the utterance of the sentence 'Uma Thurman has red eyes' is true or not. The fact that a semantic theory for a language L does not instruct L-speakers to ascertain the truth value of L-sentences is not a defect of the semantic theory. Cappelen and Lepore remind us that it is trivial that a proposition with a determinate truth value is expressed by a felicitous utterance of the sentence '100,000 years ago an insect moved over this spot', although we have no idea whether it is true or not and no idea how to find out whether it is true or not. Thinking otherwise, they say, would be to indulge in verificationism.

I find Cappelen and Lepore's response mistaken. The accusation of verificationism misses the target of our discussion. I agree that theorists, who do not adhere to verificationism, do not identify the knowledge of the proposition expressed by the utterance of a sentence with the knowledge of a method for its verification. Theorists, who are not verificationists, agree that competent speakers fully understand the proposition expressed by the utterance of the sentence

'100,000 years ago an insect moved over this spot' without being in a position to verify whether it is true or not. On the other hand, also theorists who are not verificationists cannot ignore questions as to how that sentence got the content it has and what linguistic abilities distinguish people who understand it from people who do not. Notice that I am not claiming that it is a task for semantics to find out answers to those questions. My claim is that a semantic theory must be compatible with an account that provides such answers.

A theorist, who is not a verificationist nor a semantic antirealist and takes the sentence '100,000 years ago an insect moved over this spot' to depict an epistemically inaccessible state of affairs, will not hold that the understanding of such sentence is manifested by the capacity to tell whether its truth conditions are satisfied or not. Nor can the understanding of the sentence '100,000 years ago an insect moved over this spot' be traced back to the ability to explicitly formulate the disquotational truth-condition *'100,000 years ago an insect moved over this spot' is true if and only if 100,000 years ago an insect moved over this spot (over the demonstrated spot)*, for the simple reason that many competent speakers are not able to do so. One option left is to say that a criterion for understanding is that one understands the sentence '100,000 years ago an insect moved over this spot' only if one understands the single expressions that form the sentence and the syntactic structure of the sentence. The question arises as to how the understanding of the single expressions is manifested.

It has been argued[13] that linguistic competence has two components, one inferential and the other referential. The inferential component consists in the ability to manage a network of connections among words. For example, we recognize as competent speakers those people who manifest the disposition to make the inference from, say, 'A is an insect' to 'A is an animal', or are able to give a definition of 'insect', or are able to find a synonym for 'insect', or are able to retrieve the word 'insect' from its definition, etc. The referential component consists in the ability to map words to the world. For example, the disposition to give the assent to the sentence 'that is an insect' in presence of an insect or the ability to correctly obey an order such as 'point at an insect'. This account of linguistic

[13] See Marconi 1997.

competence together with the assumption, arguably shared, that the competence in the use of the expression 'insect' requires both referential and inferential abilities demands that the following principle:

> (D) For any object *o*, 'insect' applies in English to *o* if and only if *o* is an insect.

assign the expression 'insect' a kind as its semantic content such that there must be circumstances, at least in favourable contextual conditions, in which competent speakers believe—and believe it truly—that it is instantiated. Otherwise, no matter what the linguistic competence in the use of the word 'insect' turns out to be, it is detached from the normative role of the principle (D). The result is that one gets a semantics that is useless for an overall theory of language use, since it prevents us from accounting for the linguistic practice as governed by semantic principles with normative roles.

This is the constraint that a theory of linguistic competence poses on semantics: the linguistic practice in the use of a language L needs to be taken as the manifestation of the understanding of L-expressions as governed by semantic principles with normative roles. The point of Incompleteness Arguments is that a semantic theory, which employs principles such as (A) and (B), violates such constraint. Incompleteness Arguments start with the premise that speakers are never able to believe that the property of being tall *simpliciter* or the property of being green *simpliciter* are instantiated, i.e. that the conditions for the correct application of 'tall' and 'green', as captured by the principles (A) and (B), are satisfied, because competent speakers have no beliefs about the truth value of simple sentences like 'Bradley is tall' or 'the leaves are green' independently of contextual information. Hence, the linguistic practice of competent speakers shows that their understanding of 'tall' and 'green' is not governed by the principles (A) and (B).

Analogous considerations show that learning the mastery of 'tall' and 'green' cannot amount to learning the meaning of words as governed by the principles (A) and (B). Arguably, we pick up the meaning of expressions, like 'tall' and 'green', by being exposed to assertions of simple sentences, like 'Bradley is tall' and 'the leaves are green'. Incompleteness Arguments show that assertions of simple sentences, such as 'Bradley is tall' and 'the leaves are green', cannot

be the expression of the belief that Bradley is tall *simpliciter* and the leaves are green *simpliciter*, i.e. the belief that the conditions for the application of 'tall' and 'green', as captured by the principles (A) and (B), to Bradley and to the leaves are satisfied. As a matter of fact, competent speakers have no beliefs about the truth value of those sentences independently of contextual information. If the assertions of simple sentences like 'Bradley is tall' and 'the leaves are green' are not the expression of the belief that the conditions for the application of 'tall' and 'green', as captured by the principles (A) and (B), to Bradley and to the leaves are satisfied, whatever one learns through the exposure to assertions of that kind is not a mastery of words as governed by semantic norms expressed by the principles (A) and (B).

6 Two final clarifications

The premise of an incompleteness argument registers the fact that if speakers do not take into account the contextual information, they have no beliefs about the truth value of sentences such as 'the leaves are green'. I argued that an incompleteness argument moves from that premise to the conclusion that the property of being green *simpliciter*, if any, cannot be the semantic content of the adjective 'green'. The point is semantic, not metaphysical. If speakers do not have any beliefs about when the property of being green *simpliciter* applies to objects, then they do not have any beliefs about when the conditions for the application of 'green', as captured by the axiom (B), are satisfied. This fact makes such axiom normatively idle.

One might raise the following objection. It might well be that speakers have beliefs about the truth value of sentences such as 'the leaves are green' only if they take into account the contextual information. However, this does not entail that the adjective 'green' has no invariant semantic content, i.e. a semantic content that is independent of context. One might say that whenever a speaker believes that the sentence 'the leaves are green' is true taking into account the contextual information, the speaker *ipso facto* believes that the condition for being green *simpliciter* are satisfied, and thereby the speaker believes that the condition for the application of 'green', as captured by the axiom (B), are satisfied.

An objection like this one is the obvious consequence of combining

the minimalist view in semantics with a modest metaphysical account of the property of being green *simpliciter*. The axiom (B)

> (B) For any object *o*, 'green' applies in English to *o* if and only if *o* is green (*simpliciter*)

is combined with the following modest account of what it takes to be green *simpliciter*:

> For any object *o*, *o* is green *simpliciter* if and only if *o* looks green on some surface under some circumstances.

Cappelen and Lepore do not explicitly defend such metaphysical view. They coherently refuse to be committed to it because a defense of any metaphysical theory is homework for metaphysicians and not for philosophers of language. However, they confess their sympathy to it when they respond to the following objection. Let us assume that 'the leaves are green' is true if and only if the leaves are green on some surface under some circumstances. Doesn't that make it very, indeed, too easy to be green? Doesn't that make, say, the White House green? Cappelen and Lepore respond that when we think hard about what it is to be green, maybe that is all it takes to be green. If so, then it would turn out that it is not so hard to be green. Cappelen and Lepore say that whether one finds this picture congenial or not it is not a problem that arises because of views one might hold about the context sensitivity of 'green'.

Thus, Cappelen and Lepore's response is that the above objection confuses a metaphysical issue with a semantic one. My counter-reply is that the above objection has a semantic reading. If what it takes to be green *simpliciter* is to look green on some surface under some circumstances, then any object *o* is green *simpliciter*. It follows that any sentence of the form 'O is green' is trivially true (granted the existence of O). Now, this picture is not in line with the normativity of semantic principles. An axiom such as (B) turns out to state conditions for the application of 'green' that are always trivially satisfied, because it is trivially true that anything looks green on some surface under some circumstances. This contrasts with the idea that when we learn the meaning of 'green' we learn a rule that tells us the circumstances in which it is correct to apply it apart from the circumstances in which it is not. Indeed, a consequence of semantic

minimalism combined with the above modest metaphysical view is not only that it is a trivial truth that any object is green, but also that it is a trivial truth that any object is green and red and blue and so forth for any color. In conclusion, this picture, which combines semantic minimalism with the modest metaphysical view, deprives the axiom (B) of its normative import, and I argued that this is a flaw in the field of semantics, not in the field of metaphysics.

The very same problem about normativity does not affect Contextualism in semantics, or at least some of its versions. Suppose a contextualist theory says that 'green' is a context dependent expression and its meaning is given by the rule that 'green' must be applied to an object with respect to some contextually relevant surface under some contextually relevant circumstances. Of course, selected a surface and certain circumstances in a context, it is correct to apply 'green' to an object if and only if that object looks green on that surface under those circumstances. It is not a trivial truth that an object is green in this sense. For example, it is not a trivial truth that the leaves of the Japanese maple in the photographer's studio have been painted green.

I dedicate a final reflection on the argument for the existence of invariant contents that says that although they do not fit speakers' intuitive judgments about the truth conditional content of assertions in contexts, they nevertheless play an indispensable role in communication and, contrary to what some contextualists hold, they are psychologically real.

Cappelen and Lepore maintain that invariant *semantic* contents play a function in the cognitive life of communicators that no other content can play.[14] The idea is that invariant contents have a role to play as fallback content, i.e. the content which is guaranteed to be recoverable in a communicative exchange when something goes wrong due to the fact that either the speaker or the hearer or both have an imperfect, partial, limited, erroneous grasp of the contextual information. The invariant content is that content the speaker (the audience) can expect the audience (the speaker) to grasp (and expect the audience (the speaker) to expect the speaker (the audience) to expect them to grasp) even if they have mistaken or incomplete contextual

[14] See also Borg 2007, 2009 and 2012.

information. Cappelen and Lepore say that the invariant content is our defense against confusion, misunderstanding and mistakes. Even if the invariant content is trivially true, as in the case of an object being green on some surface under some circumstances, nonetheless it is a starting point from which the content that the speaker intended to communicate can be recovered. Therefore, the invariant contents are psychologically real.

My reply is that this argument is a *non sequitur*. The conclusion that certain expressions are not context dependent and have an invariant *semantic* content does not follow from the premise that invariant contents play an important role in communication. Consider the expression 'I'. Nobody will put in question that 'I' is a context dependent expression. 'I' is an indexical which Cappelen and Lepore put into the basic set of context dependent expressions. Suppose one overhears the utterance of the sentence 'I have headache' coming from the next room without having access to the contextual information, i.e. without knowing who is the speaker of the utterance. This is a case in which something goes wrong due to the fact that one has an imperfect grasp of the contextual information. Nonetheless, there is a content that one can understand in virtue of being a competent speaker. One understands that the speaker of that utterance has headache. That there is a unique speaker of that utterance who suffers from headache is a content that one can grasp even if one does not know who is the speaker, and therefore one cannot grasp what the speaker said, i.e. the *semantic* content of that utterance. The recovered content might play an important role. One can enter the next room and ask who uttered the sentence 'I have headache' in order to discover who is the speaker, and hence in order to understand the content that the speaker semantically expressed. Although there is an invariant content that one can grasp in virtue of being a competent speaker, it does not follow that the expression 'I' is not context dependent. In general, as far as indexicals and demonstratives are concerned, competent speakers can recover invariant contents from their characters, and such contents can play an important role in communication. Of course, it does not follow that indexicals and demonstratives are not context dependent expressions.

Consider now a contextualist theory that says that 'green' is a context dependent expression with the rule that 'green' applies to

an object with respect to a contextually relevant surface under some contextually relevant circumstances. Suppose one overhears an utterance of the sentence 'that is green' coming from the photographer's studio without having access to the studio. In virtue of being a competent speaker, one knows that what the speaker said is true if and only if there is a contextually relevant object that has a contextually relevant surface looking green under some contextually relevant circumstances. This is not what the speaker said. The speaker semantically expressed the proposition that that Japanese maple has the leaves repainted green. Although one cannot grasp such proposition, which is the *semantic* content of the assertion, the *recovered* content one understands is a starting point that might lead to grasp the semantic content.

Thus, my conclusion is that no doubt there are invariant contents that can be associated with certain expressions in virtue of being recoverable from our knowledge of their meaning. No doubt such contents are psychologically real and might play important roles in communication. However, it does not follow that those expressions are not context dependent.

7 Conclusions

I argued that the conclusions of Incompleteness Arguments are not that certain entities do not exist. Those are metaphysical questions that metaphysicians are called to answer. Contrary to Cappelen and Lepore's view, and no matter what metaphysicians are willing to say, Incompleteness Arguments show that even if one acknowledges the existence of certain entities, e.g. the property of being tall *simpliciter* and the property of being green *simpliciter*, such entities cannot be the contents that a semantic theory associates with words, because a semantic theory so construed is incompatible with theoretical considerations about language learning and language understanding.

One can agree with Cappelen and Lepore on keeping issues in metaphysics apart from issues in the philosophy of language and on rejecting Dummett's thesis (I). One can also agree with Cappelen and Lepore on rejecting Dummett's thesis (II) and its constitutive constraint that the linguistic competence must *constitute* the implicit knowledge of semantics, which, in Dummett's view, is the premise

that leads to semantic antirealism. However, one cannot go too far, as Cappelen and Lepore go, in detaching semantics from linguistic competence. There is a constraint that a theory of linguistic competence poses on semantics: the linguistic practice needs to be taken as the manifestation of the understanding of words and as the basis for the learning of their meaning *insofar as* they are words governed by semantic principles with normative import. If certain semantic principles are not suitable for such an account of linguistic competence, then any semantic theory that endorses them is on the wrong track. I take this result, which points at an intimate connection between linguistic competence and semantics, to be an important part of Dummett's legacy in the philosophy of language.

Massimiliano Vignolo
Dept. of Philosophy
University of Genoa
16124 Genoa, Italy
maxi@nous.unige.it

References

Borg, Emma. 2007. Minimalism versus Contextualism in Semantics. In *Context-Sensitivity and Semantic Minimalism*, ed. by G. Preyer and G. Peter. Oxford: Oxford University Press, 339-359.

Borg, Emma. 2009. Minimal Semantics and the Nature of Psychological Evidence. In *New Waves in Philosophy of Language*, ed. by S. Sawyer. Basingstoke: Palgrave Macmillan, 24-40.

Borg, Emma. 2012. Semantics without Pragmatics. In *The Cambridge Handbook of Pragmatics,* ed. by K. Allen and K. Jaszczolt. Cambridge: Cambridge University Press, 513-528.

Cappelen, Herman and Lepore, Ernest. 2005. *Insensitive Semantics. A Defense of Semantic Minimalism and Speech Act Pluralism*, Oxford: Blackwell.

Carston, Robyn. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

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

Devitt, Michael. 2007. Referential Descriptions: a Note on Back. *European Journal of Analytic Philosophy* 3: 49-53.

Devitt, Michael and Sterelny, Kim. 1999. *Language and Reality*, 2nd edition. Oxford: Clarendon Press.

Dummett, Michael. 1975. What is a Theory of Meaning? In *Mind and Language*, ed. by S. D. Guttenplan. Oxford: Oxford University Press, 95-138.

Dummett, Michael. 1976. What is a Theory of Meaning? (II) In *Truth and Meaning*, ed. by G. Evans and J. McDowell. Oxford: Clarendon Press, 67-137.
Dummett, Michael. 1977. *Elements of Intuitionism*. Oxford: Clarendon Press.
Dummett, Michael. 1978. *Truth and Other Enigmas*. London: Duckworth.
Dummett, Michael. 1981. *Frege, Philosophy of Language*, 2nd edition. London: Duckworth.
Dummett, Michael. 1991. *The Logical Basis of Metaphysics*. London: Duckworth.
Marconi, Diego. 1997. *Lexical Competence*. Cambridge MA: MIT Press.
Putnam, Hilary. 1975. The Meaning of 'Meaning'. In *Philosophical Papers* vol. 2, *Mind Language and Reality*, Cambridge MA: Cambridge University Press, 215-271.
Recanati. Francois. 2011. *Truth-Conditional Pragmatics*. Oxford: Clarendon Press.
Stanley, Jason. 2007. *Language in Context*. Oxford: Oxford University Press.
Taylor, Kennet. 2001. Sex, Breakfast, and Descriptus Interruptus. *Synthese* 128: 45-61.
Tennant, Neil. 1987. *Antirealism and Logic*. Oxford: Clarendon Press.
Travis, Charles. 1997. Pragmatics. In *A Companion to the Philosophy of Language*, ed. by B. Hale and C. Wright. Oxford: Blackwell, 87-107.
Travis, Charles. 2008. *Occasion-Sensitivity*. Oxford: Oxford University Press.
Wright, Crispin. 1983. *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.

The Problem of Nonexistence: Truthmaking or Semantics? Critical Notice of *The Objects of Thought*, by Tim Crane

Lee Walters
University of Southampton

BIBLID [0873-626X (2015) 41; pp. 231-245]

Tim Crane's *The Objects of Thought* is, I think, a much needed corrective to standard ways that analytic philosophers think about nonexistence. It starts from our common sense thought and talk, and tries to carve out a position that can defend this starting point in the face of criticism. It is well-written, a pleasure to read, and largely clear. I would recommend it to anyone interested in the problems of nonexistence. In §1 I sketch Crane's central ideas about the nonexistent, before turning to themes that I would like to have heard more about. In §2, I distinguish two problems of nonexistence, showing that whilst Crane solves one, he does not address the other. Although Crane did not seek to address both problems, I think we should recognize that there is this residual problem of nonexistence remaining. Next (§3), I argue that whilst Crane is correct to think that a negative free logic has to be rejected if we construe it as making a claim about grammatical subject-predicate sentences, we might be able to salvage it if we recognise a class of logical predicates. But whether this is possible or not, depends on the solution to the unaddressed problem of nonexistence. In the final two sections I briefly raise a concern about Crane's view of quantification, before making a suggestion about how his view might be employed in addressing Geach's problem of intentional identity.

1 Crane's approach

Some of the things we think about exist, like Buda Castle, but some

Disputatio, Vol. VII, No. 41, November 2015
Received: 19/11/2015

of the things we think about, like Hogwarts, do not. Hence there are truths about the nonexistent, such as Lee is thinking about Hogwarts. And yet the world does not contain more than what exists. Tim Crane's task is to defend and reconcile these apparently conflicting common sense claims.

But wait, you might think. Do we really think *about* the nonexistent? Surely, if I am thinking about something, there must really be something for my thought to be about. Thinking about is, so the objection goes, what Crane calls a 'real' or 'substantial' relation, and so entails its relata. Certainly this suggestion is not without merit, and even those friendly to providing accounts of empty names, such as Mark Sainsbury (2005: 237-238), have denied that we do think about the nonexistent. Still, Crane is correct that our thoughts are characterized in certain ways, even when there is nothing in reality that we are thinking of, and it seems that the English word 'about' is as good a way as any to capture this phenomena. So, just as there can be drawings and sculptures *of* Peter Pan, there can be thoughts *about* him too. Crane does, however, recognize that there is a real relation in area, and he reserves the word 'reference' for this relation: one can think or talk about Peter Pan, but one cannot refer to him.

I am thinking about Peter Pan; Peter Pan does not exist; therefore, *some* of the things I am thinking about do not exist. So as well as thinking and talking about the nonexistent, we can also quantify "over" them. It seems to be part of the data that some of the things we think about do not exist. Moreover, for Crane, so-called existential readings of 'there is' sentences are semantically equivalent to their corresponding 'some' sentences. This is because 'there' is a semantically vacuous term, present simply because syntax requires it. So, *there are* things we think about that do not exist. But this is not an ontological claim, since, for Crane, nonexistents are no part of reality in any sense. Rather, it "is simply another way of saying that we can genuinely think about things that don't exist" (2003: 5).

Crane's view is, then, something of a hybrid. It resembles a positive free logic in that it allows for true seemingly simple sentences containing non-referring terms, but it is Meinongian in that its unrestricted quantifiers range "over" nonexistents: in free logics traditionally conceived, the quantifiers range over only existents, and traditional Meinongian pictures allow for reference to nonexistents.

In this way, Crane's picture is an improvement on these rival views. It seems to be a fact about English that there are true generalizations about nonexistents, and I myself cannot make sense of a picture on which we refer to, as opposed to talk about, the nonexistent, since there are not really any.

Crane, following standard usage, reserves the symbols ∃ and ∀ for quantifying over the existents. Because Crane allows for meaningful empty names, these quantifiers are subject to a free logic, and so the rules for existential generalization and universal instantiation need to be restricted to cases in which ∃x (x=a). But as we have seen, Crane does not think the English word 'some' corresponds to ∃. Rather than 'some' expressing the existential quantifier, Crane could have followed others in saying that it expresses a particular quantifier, Σ, and that 'all' has a corresponding reading, Π, that ranges "over" both nonexistents and existents. These quantifiers behave classically with unrestricted particular generalization and universal instantiation, and so allow for the move from 'I am thinking about Peter Pan' to 'Σx Lee is thinking about x'.

Despite sharing some features with other forms of neo-Meinongianism, Crane's view differs sharply in that he rejects any form of characterization principle along the following lines

> CP: Nonexistents are the way they are characterized as being by the appropriate myth, theory, fiction, etc.

Moreover, he differs from some positive free logicians in that he denies that nonexistents can have any ordinary properties such as being a horse, being a detective, or being located in space. For Crane, these are existence entailing properties, and so cannot be had by nonexistents.

So what truths concerning the nonexistent does Crane allow for? For Crane these fall in to three categories. First, there are negative existential claims such as Hogwarts does not exist. Second, there are representation-dependent truths, examples include being thought about, being famous, and being a fictional character. Third, Crane allows for the truth of trivial identity statements, such as Peter Pan is Peter Pan (although how to spell-out what a trivial identity statement is is not itself trivial (2013: 165)). Crane's task is to provide an account of how these statements about the nonexistent can be true

given that the nonexistent are no part of reality.

Crane offers a "metaphysical reduction" of these claims about the nonexistent, providing truth makers for these truths, in his non-committal, not theoretically-loaded sense of this phrase. But unlike neo-Meinongians, he wants to do this without metaphysical extravagance, and so eschews appeals to Meinongian object theory with its reliance on impossible worlds, the distinction between nuclear and extra-nuclear properties, or between encoding and instantiating. So how exactly does Crane account for the truths above?

First, nonexistence claims are made true simply in virtue of the fact that the world does not contain nonexistents. As Crane puts it, "the falsity of 'Vulcan exists' is ensured by the fact that reality ... does not contain Vulcan" (2013: 119); there is no truth maker for 'Vulcan exists'. So given that 'Vulcan exists' is false, its negation is true. And as Crane notes (2013: 73), this negation is expressed by 'Vulcan does not exist'.

Second, representation-dependent truths are true, as the name suggests, in virtue of the existence of some representation, whether this is a story, a theory, or an episode of thinking. So, for instance, 'Vulcan was a planet postulated by Le Verrier' is true iff there was an event where Le Verrier represented Vulcan as a planet in certain way (2013: 135).

Third, self-identity claims follow from the "logical truth that for all x, x=x" (2013: 165), where this must be understood as $\Pi x\ x=x$, if it is to yield, say, that Vulcan=Vulcan by universal instantiation. But why think $\Pi x\ x=x$ is a logical truth? Well, in classical systems it follows from the rule for introducing identity, a=a, by universal introduction. But whether a=a is a logical truth is precisely what is at issue, so Crane cannot appeal to this to justify his claim. So I think that Crane has not provided any independent reason for us to accept that nonexistents are in fact self-identical. Moreover, there does not seem to be anything in the world to ground these self-identity claims; 'Vulcan=Vulcan' has no truth maker as Crane admits (2013: 163). I think it would be simpler and more in keeping with Crane's project to deny that these claims are in fact true.

Apart from the three types of truth about the nonexistent that Crane explicitly discusses, we should also count as true negative claims made with any existence entailing properties, not just negative

existentials: if it is true that Vulcan does not exist, given the absence of Vulcan, then it is also true than Vulcan is not a planet, since being a planet is existence entailing. Finally, Crane might want to consider which modal claims concerning nonexistents are true. Perhaps Vulcan could not have existed? If so, perhaps some modal claims are also existence entailing, and so can be subsumed by the previous point.

Despite what many philosophers have said about quantification and empty names, Crane's general picture above seems dead right to me. It is, I think, on the basis of theoretical considerations that have not been adequately justified, that some resist this intuitive picture. So Crane is to be applauded for spelling-out this common sense picture of the nonexistent, and rejecting philosophical orthodoxy. And yet, some will not be completely satisfied with Crane's solution.

2 The problems of nonexistence

Crane states the problem of nonexistence as follows: "if truth is supervenient on being, then how can one truly say *of something that is not*—something that has no being—that it *is* a certain way? How can such a claim be true?" (2003: 20). It is not entirely clear what the tension is supposed to be here. After all, does anyone think that there are possible worlds where the existence facts are the same, but the truths about nonexistents differ? Supervenience is not really the issue here, I think. For Crane, the issue is better put in terms of truth making. The problem of nonexistence, as Crane thinks of it, is that given that nonexistents are not a part of reality, what are the truth makers for statements about them? As Crane himself puts it, "Given that when something is true, it is reality that makes it so, we are obliged to ask: what in reality makes these claims about the non-existent true?" (2013: 118). As we saw above, Crane sets out to answer this question, and his answer seems on the right lines to me.

But is this enough? Crane describes his reductionism as providing an explanation of the truth of statements about the nonexistent without giving the meaning of those statements. We might, however, also want an account of the meaning of such statements. The residual problem of the nonexistent, unaddressed by Crane, is how to provide systematic truth conditions for claims about the nonexistent, without appealing to reference to nonexistents.

Now Crane can be forgiven for not engaging in this no doubt difficult, and largely technical semantic project. It is fine for there to be a division of philosophical labour, and Crane's positive picture was well worth setting-out as a much needed alternative to more extravagant approaches. Still, there is this residual problem, and until this problem has been solved, Crane's extravagant opponents, at least, will view his approach with suspicion.

A related worry comes from asking how Crane thinks we should decide ontological questions? He says that what people are in fact committed to is a matter of what they believe in, rather than what they quantify over. Fair enough. But we can ask what *ought* they, objectively, believe in. How do we settle that question? Crane does not think that there is an informative formal criterion of which object-language sentences are ontologically committing (see below). So perhaps he thinks ontological commitment is determined by which entities are appealed to in the metalanguage when giving the semantics of the object-language. But Crane, as we have seen, does not provide such a semantics. His opponents may suspect that once he does provide a semantics he will find himself faced with the same extravagant choices he criticizes.

As well as Crane's truth maker conception of the problem of nonexistence, then, there is another problem that we can characterize with the following inconsistent triad (I do not say these two problems exhaust the problems of nonexistence):

(1) There are true subject-predicate sentences about nonexistents.

(2) If a subject-predicate sentence '*a* is F' is true, then '*a*' refers.

(3) At least one subject term in a subject-predicate sentence about nonexistents lacks a referent.

Crane, effectively, takes (3) to be a constraint on the solution, and I agree. He also takes (1) to be constitutive of the problem, so, he rejects (2). But (2) follows from the simple view of truth

SVT: A predicative sentence, '*a* is F' is true iff the object denoted by '*a*' has the property ascribed by 'F'.

And so as Crane rejects (2), he also rejects SVT. But he does not provide a systematic alternative to SVT, which leaves open the questions just raised. In the next section, I sketch some thoughts on Crane's account of properties and predicates, and a different way of thinking that is nonetheless consonant with his whole approach.

3 Properties and predicates

Crane claims that there are true subject-predicate sentences about the nonexistent. Moreover, he says that these sentences are true because the nonexistent the sentence is about has the property ascribed by the predicate. Does this, then, not allow him to answer the challenge faced above? As Crane notes, on his view "The truth-conditions for a claim of the form '*a* is *F*' is that it is true just in case *a* has the property *F*. We can state the truth-conditions in this form, in the same way, whether or not '*a*' refers to anything" (2013: 58).

Here it might look like Crane is going some way to providing the systematic theory I asked for. This impression is, I think, illusory (not that Crane claims otherwise) since the properties that nonexistents have are, for Crane, "pleonastic", the result of the grammatical transformations from '*a* is F', to 'there is a property that *a* has, namely Fness'. As a result, to say that '*a* is *F*' is true just in case *a* has the property *F* is not to provide an *explanation* of why '*a* is F' is true. They are simply two ways of saying the same thing.

Now one way of holding on to SVT, but allowing for truths about the nonexistent, is to adopt a negative free logic that supplements SVT with

> NFL: If '*a*' does not refer, then any subject-predicate sentence, '*a* is F' is false.

With NFL we can account for the falsity of all of the existence entailing claims concerning nonexistents, and thus for the truth of their negations, including negative existentials. But although NFL is consistent with their being truths about nonexistents, it does not allow for true subject-predicate claims about the nonexistent, and so it resolves the residual problem of nonexistence by rejecting (1). Although Crane would be happy to accept this approach for a range of sentences, he rejects it in its full generality, since he thinks that

it cannot provide a satisfactory account of representation-dependent truths. In brief, this is because Crane thinks that not all of these truths can be accounted for by employing intensional operators taking subject-predicate sentences within their scope. Rather, Crane thinks that there are true representation-dependent, subject-predicate sentences concerning nonexistents, and so NFL has to be rejected.

Crane thinks that there are true predications concerning nonexistents because he follows Dummett (1973: 37-38) in saying that a predicate is what results when we remove one or more referring expressions from a sentence. There are at least three worries that we might raise for this conception. First, one might want to exclude certain complex sentences from this method of predicate formation, otherwise we can have what appear to be incompatible predicates true of the same object. For instance, Frege's puzzle might give rise to the predicates 'Lee believes that x is F' and 'Lee does not believe that x is F' (as opposed to 'Lee believes x is not F'). Second, even ignoring complex sentences, this method might be objected to because it allows for a predicate 'Professor x was an expert on Tarot' to be generated by removing 'Dummett' from 'Professor Dummett was an expert on tarot'. But it does not make any sense to predicate this of an object, as can be seen by completing the predicate with some other phrase that picks out Dummett, such as 'the Wykeham Professor of Logic in 1985'. This problem could be avoided, however, by placing a suitable restriction on what counts as a referring term in the relevant sense. But even leaving all this to one side, there is a third problem which is brought out by considering Quine's (1960: 153) example of

(4) Giorgione was so-called because of his size.

By the Dummett method of predicate formation, this yields the predicate

(5) x was so-called because of his size.

But it is odd to say that (5) is a predicate. First, it does not allow for substitution of co-referring terms, even when we concern ourselves with de re readings. For it is not true that

(6) Barbarelli is such that he was so-called because of his size.

Relatedly, one cannot quantify into this predicate since neither

(7) ∃x (x was so-called because of his size)

(8) Σx (x was so-called because of his size)

make sense. But it seems to me that the notion of a predicate is tied as much to quantification as it is to combining with singular terms.

The real predicate involved in (4) is more perspicuously given by

(9) x was called 'Giorgione' because of his size

and (9) is not subject to the problems above.

Moreover, (4) puts pressure on the notion of a pleonastic property, since we cannot move from (4) to

(10) There is a property Giorgione has, namely so-called because of his sizeness.

What this shows, then, is that it is not as harmless as Crane suggests to think of true claims concerning nonexistents as true subject-predicate claims where the nonexistent has a pleonastic property corresponding to the predicate. The point of this is to bring out that as well as Dummett's grammatical notion, we also have the separate notion of a logical predicate. And it is the logical notion, I suggest, that is important to the assessment of NFL. Crane rejects NFL because

> The mere idea of a sentence free of truth-functional operators, and of 'intensional' operators ... is clear enough, but [examples like 'Vulcan was a planet postulated by Le Verrier'] show that these restrictions do not on their own determine a kind of expression which always determines a falsehood when combined with a non-referring term. There does not seem to be a syntactic or formal criterion of simplicity [for a predicate in NFL's sense] (2013: 55).

Now I think the negative free logician has more formal resources than Crane considers. For one thing, the passive versions of representation-dependent truths often sound much worse than the active forms: compare 'Le Verrier is thinking about Vulcan' with 'Vulcan is being thought about by Le Verrier'. But not all do, and it is not clear

that there is surface-form syntactic criterion of a logical predicate. Still, it might be correct that there is an interesting class of expressions, logical predicates, that when combined with an empty name, always produce a falsehood, but that this class cannot be read-off surface structure. The only way to discover if this is true is by doing the semantics and discovering the logical forms of the problematic sentences. If there is a class of logical predicates that combine with empty names to produce false sentences, then perhaps the thought behind NFL is vindicated. Further, these logical forms would correspond to the existence entailing properties, and so we would have an explanation of which properties are existence entailing.

What about the representation-dependent truths? If these are not logical predications, what are they? It seems to me that what (many of) these truths are doing is not ascribing a property to, or predicating something of a nonexistent, in some intuitive sense that has not been made precise. Rather, they are *characterizing* representations as Peter Pan-sculptures, Vulcan-theories, Holmes-stories, Pegasus-thoughts, etc. Whether, ultimately, this approach can be sustained to defend NFL is not clear, but it is only by investigating the logical forms of sentences that we can find out. In any case, this approach, which takes *characterizing* as primitive (see Forbes 2006) seems to fit well with Crane's (2013: 90) proposal to take intentionality as primitive.

Regardless of the logical forms of claims about the nonexistent, Crane is right to reject NFL as a claim about (surface) syntactically simple sentences. But by investigating why it is false read as such, by seeking to provide a systematic semantics for the nonexistent, we open up the possibility of drawing some worthwhile logical distinctions between sentences that are genuinely ascribing properties of their subjects and those that are not, and between claims that entail the existence of their subjects, and those that do not. Consequently, we might be able to ward off the suspicions of some of Crane's opponents, and maintain the possibility of doing ontology in something like the Fregean tradition. All of this goes beyond what Crane sought to do in his book. And as I have said, I think that his general picture and metaphysical reduction must be correct. Still, I think some investigation of these issues would have been interesting.

4 Quantification

Crane, as we saw allows for quantification "over" nonexistents. I have repeatedly used scare quotes because it was not entirely clear to me what Crane's account amounts to exactly. Crane says that he wants to keep "the basic ideas of the logic of quantification intact" (2013: 31). So what, then, does it mean to quantify "over" nonexistents for Crane?

> It is to have non-existent objects of thought in the universe of discourse, where ... to have an object of thought in the universe of discourse is to have it among the things relevant to what we are talking about ... These things can be 'values' of the variables bound by the quantifiers, just in the sense that things can be true or false of the objects of thought. So, when evaluating 'some biblical characters did not exist' we look for something in the domain (biblical characters) of which we can predicate non-existence. And lo! We find one: Abraham. Abraham is then a value of the variable (2013: 40).

Note how Crane himself uses scare quotes for 'values'. If Crane wants to say that we can quantify over nonexistents in the way in which standard logic quantifiers over a domain of existents, then I would like to have seen more detail about assignment functions, satisfaction, and the like to help me fully understand what was going on. But it seems to me that Crane does not need to go down this route, since, at other points, his account of quantification does not appear to amount to quantifying over nonexistents. Rather, it seems to be a device of generalizing into certain syntactic positions:

> After all, if we can use a name to talk of something which does not exist, then the quantifier 'some' is just a generalization from the use of a name (2013: 16).

> quantified sentences [such as] ... 'Some characters in the Bible did not exist'—are best understood as generalizations from sentences that predicate something of their subjects (2013: 119).

I would have liked to have heard more about whether this kind of syntactic generalization was what Crane had in mind, and also how his approach compares with others who have adopted such approaches.

5 De re thought

After setting out his metaphysical picture, Crane (2013, chapter 6) turns to the problem of thinking about specific nonexistents. There is too much in this chapter to cover, so I shall just focus on his discussion of de re thought. Here, as is standard, Crane construes the de re/de dicto distinction syntactically, so that quantifying into a belief report, say, counts as de re.

After noting that singular thoughts can be attributed de dicto, Crane considers whether singular thought entails a de re attribution. Crane notes that whereas on the orthodox conception, this is true, since 'S believes that ... a ...' entails '$\exists$x (S believes ... of x)', no such entailment is forthcoming on Crane's account, since we can believe things about the nonexistent. Rather than take this as counting against singular thought about the nonexistent, Crane instead rejects the idea that singular thought entails de re attribution.

Now clearly Crane is correct that beliefs about the nonexistent do not license *existential* generalization. And so if existential generalization is required for the de re, then singular thought about the nonexistent does not entail a de re reading. But why think that *existential* quantification is required? The syntactic construal of the de re does not mention existential quantification. Moreover, given that Crane employs something like a particular quantifier that quantifies "over" nonexistents, he is free to acknowledge de re attributions of belief concerning nonexistents. For example, Σx such that Crane believes x does not exist.

Two options present themselves. First, Crane could accept that singular thought, even about the nonexistent, does entail a de re attribution, albeit one in terms of the particular, rather than the existential, quantifier. Second, he could reject the syntactic criterion of the de re given above, in favour of a relational construal of the de re. This seems to fit better with Crane's way of thinking since he glosses 'de re' at several points as 'relational', where I take him to mean substantially relational. If this is right, then there is no purely syntactic characterization of the de re for Crane, just as there is no syntactic construal of the ontologically committing claims.

But having pulled apart the syntactic and relational construals of 'de re', it seems as if Crane is in a position to provide an irrealist

construal of the problematic Geach sentence concerning intentional identity:

(11) Hob thinks a witch blighted Bob's mare, and Nob wonders whether she (the same witch) killed Cob's sow.

It has been thought that (11) cries out for a syntactically de re reading. But on the standard assumption that quantification is ontologically committing, such a reading commits to there being something in reality that Hob and Nob's mental states are about. But such a consequence is unwelcome. However, once we sever the link between quantification and ontological commitment, as Crane does, we can give a syntactically de re reading without these unwanted consequences along the lines of the following:

(12) Σx (x is a witch) such that Hob thinks that (x is a witch and) x blighted Bob's mare, and Σy such that Nob wonders whether y (a witch) killed Cob's sow, and x $\approx$ y,

where the material in parentheses can be included or not depending on how exactly we read (11). Two comments. First, 'x $\approx$ y' means x is the same as y, to be discussed below. Second, as Nathan Salmon (2015) notes, it seems plausible to suggest that 'witch' has a reading on which it can be truly predicated of mythical witches, and also a reading which means something like 'is a mythical witch or a real witch' (compare 'gun' and 'poet', in 'is that gun real or fake' and 'how many poets are there living or buried in Budapest?' (cf. Partee 2010). If so, Crane can take the occurrence of 'witch' outside the scope of the propositional attitudes as not committing to real witches.

But what of 'x $\approx$ y'? Aside from the trivial identity statements discussed above, Crane does not allow for true identity statements concerning nonexistents, so x $\approx$ y cannot be treated as x=y. Crane (2013: 163-164) suggests that we cash out 'x $\approx$ y' in terms of the resemblance of representations: Mercury and Hermes are not literally identical, but we can say that they are "the same", by virtue of the similarity of the representations of x and y. For some purposes this might be right, but I think that often our sameness talk reflects more than qualitative similarity. If I say that you and I have the same car, what this would ordinarily mean is that we have the same *type*

of car, such as a VW Golf. But being a VW Golf is not (merely) a matter of resemblance, for causal links are important too—if your car just happens to look like a VW Golf then it is not in fact a VW Golf. How however we cash out this talk of types, we cannot employ the same treatment in the case of nonexistents: nonexistents do not fall under any causally-individuated type, since they don't exist. Nevertheless, I think that to account for some of our sameness talk concerning the nonexistent, we must appeal to causation, since it seems that whether we count fictional characters as being the same (from an irrealist perspective) depends on whether the uses of the names we use to speak of them are related. If so, it might be helpful for Crane to appeal to Sainsbury's (2005, chapter 3) name-using practices, and then to ground (some) sameness talk in terms of causally related name-using practices along the lines of Salis (2013). But as long as Crane has a satisfactory account of 'x ≈ y', it seems as if he might be well-placed to offer an account of (11).[1]

Lee Walters
Department of Philosophy
University of Southampton
Avenue Campus
Southampton, S017 1BF
l.walters@soton.ac.uk

References

Crane, T. 2013. *The Objects of Thought*. Oxford: Oxford University Press.

Dummett, M. 1981. *Frege Philosophy of Language*. Second Edition. London: Duckworth.

Forbes, G. 2006. *Attitude Problems*. Oxford: Clarendon Press.

Partee, B. 2010. Privative Adjectives: subsective plus coercion. In *Presuppositions and Discourse: Essays offered to Hans Kamp*, ed. by Rainer Bäuerle, Uwe Reyle and Thomas Ede Zimmermann. Bingley, UK: Emerald Group Publishing, 273-285.

Quine, W.V.O. 1960. *Word and Object*. Cambridge, MA: MIT Press.

Sainsbury, R.M. 2005. *Reference without Referents*. Oxford: Oxford University Press.

[1] Thanks to the Institute of Advanced Study at the Central European University for a Junior Fellowship during which I wrote this piece.

Salis, F. 2013. Fictional Names and the Problem of Intersubjective Identification. *Dialectica* 67: 283-301.
Salmon, N. 2015. The Philosopher's Stone and Other Mythical Objects. In *Fictional Objects*, ed. by Stuart Brock and Anthony Everett. Oxford: Oxford University Press, 114-128.

Book Reviews

BIBLID [0873-626X (2015) 41; pp. 247-260]

Born Free and Equal? A Philosophical Inquiry into the Nature of Discrimination, by K. Lippert-Rasmussen. New York, NY: Oxford University Press, 2014, xii + 317 pages, ISBN 978019 9796113.

Kasper Lippert-Rasmussen begins his recently published book on discrimination by distinguishing three main general questions that are undertaken in the book, and that organize its structure, namely: what discrimination is, what makes it wrong, and in which cases differential treatment is discriminatory, or what should be done about wrongful discrimination. Both the approach and the layout of these questions make the book a thought-provoking rewarding reading.

In part 1 of the book, Lippert-Rasmussen examines several types of discrimination. This analysis is not an exhaustive taxonomy, but one that allows the reader both to identify a reference framework to allocate the moral wrongness of discrimination, and gives a glimpse of a proposed counteraction of discrimination acts, and its consequences. By doing so, the author advances the content and motivation of parts 2, and 3 of the book. In part 1, Lippert-Rasmussen advances many highly relevant debates on discrimination. Among the debates presented in this part, high points that merit further discussion include: what he calls the *generic definition* of discrimination which offers a broad, in contrast to the usual narrow definition, approach to discrimination. To wit, Lippert-Rasmussen defines *generic* discrimination as follows: "to discriminate against someone is to treat her disadvantageously relative to others because she has or is believed to have some particular feature that those others do not have." Another relevant point is the decision to stick to *group discrimination* as the approach that better accounts both what generally bothers people of discrimination, and the detailed treatment of indirect discrimination. On the one hand, one of the reasons he offers for this move is that, in his view, most of the times when something is said

to be discriminated, it *concerns* group discrimination. On the other hand, indirect discrimination is understood in his account as an example of non-intentional discrimination.

Particularly interesting here is the debate he opens on the *generic* definition of discrimination. In a nutshell, discrimination is defined as disadvantageous differential treatment. Far from a discrimination skeptic that disregards affirmative action, he aims at revisiting the concept of discrimination, and what is morally wrong about it from the very beginning. This intention is clear when he argues in page 15 that: "there is not even a presumption that someone who engages in *generic* (italics added) discrimination acts wrongly." However, the parts of the book that analyze cases of what we may call advantageous differential treatment, or non-wrongful discrimination, are somewhat unclear. For example, in pages 23, 25, and 27, Lippert-Rasmussen argues that nepotism is not a discriminatory act in the relevant sense, while in pages 41 to 46 what qualifies as advantageous differential treatment remains vague.

Lippert-Rasmussen moves a step forward in the definition of discrimination and states in page 16 that discrimination is *essentially comparative* with respect to individuals. The author believes that a feature that may turn generic discrimination to be morally wrong, or at least morally relevant, lies in (unjustified) disadvantageous treatment in comparison to others. This further feature of *generic* discrimination opens the floor to make a relevant distinction. Whilst he states that equal treatment and even non-disadvantageous discrimination may well not be morally wrong, he also considers that compared differential treatment between two people is morally wrong. A more in depth discussion of what makes discrimination wrong is undertaken in part 2 of the book. The relevant differential background of both conceptions is that while the *comparative* account identifies the moral wrongness of discrimination as due to the inequality that it generates, the other account perceives the wrongness in the discriminatory act. In the latter sense, the wrongness would not be only based on the effects generated neither on a particular situation, nor on further counterfactual situations, but on the unjustified differential treatment to a member of (in Lippert-Rasmussen's account) a social salient group. To illustrate the point: to assess whether someone is discriminating another in a morally relevant way, it should

be first established how that person would treat a subject from another equally salient social group in the same situation. It may be said that what justifies this comparison remains somewhat unclear. This feature of the author's account of the wrongness of discrimination defines his characterization of the harm-based account, and in particular of harm as the necessary condition of the wrongness of discrimination, most of all in pages 160 and 161.

In chapter 1, Lippert-Rasmussen sticks to *group discrimination* as a descriptive concept that, in his view, better explains what people talk about when they talk about discrimination. He points out that although *group* discrimination is the proxy for an account of the wrongness of discrimination (and many times in the text it seems to be the only objectionable type of discrimination), it just is a necessary condition for wrongful discrimination, but not a sufficient one. In other words, group discrimination is not always morally wrong. Lippert-Rasmussen proceeds to distinguish different senses in which it might be morally wrong. The reasons given in favor of establishing group discrimination as the main concept that qualifies as discrimination finishes at this point. Although Lippert-Rasmussen's view in this point is not clear, the reader may intuitively guess that he remains neutral on the distinctions made regarding the wrongness of group discriminatory treatment. To wit: Lippert-Rasmussen remains neutral about the moral distinction between direct and indirect discrimination, cognitive and non-cognitive discrimination, and valuation based and non-valuation based discrimination. A clear position regarding these subjects would have been helpful to clarify some normative points in part 3 of the book. It would also have been helpful to have a clear characterization of when is (unjustified) disadvantageous treatment morally wrong.

In part 2 of the book, Lippert-Rasmussen assesses three concrete accounts of the wrongness of discrimination: Larry Alexander's account on objectionable mental states, conditioned by false believes, and resulting in bias; Deborah Hellman's account on discrimination demeaning equal human worth; and Thomas Scanlon's account on the offensive meaning of discrimination. High points include: intrinsically wrong discrimination, instrumental reasons to assess the moral wrongness of discrimination, objective meaning accounts, Lippert-Rasmussen's harm-based account of the wrongness of dis-

crimination, and his version of a prioritarian harm-based account—a desert-prioritarian account. Briefly, though not less relevant, it should be noticed that one of the main difficulties for Lippert-Rasmussen's desert-prioritarian account is the prioritarian calculus. According to the desert-prioritarian account, individuals which are comparatively worse have greater moral value than those that are comparatively better off. While Lippert-Rasmussen is aware of some objections regarding equal value of both the discriminatee, and the discriminator (166), and accommodates some cases to his account, the metric of prioritarian calculations remains unclear.

Particularly interesting here is his approach to harm-based accounts of discrimination (154 ff) to which the author is more sympathetic. Broadly, Lippert-Rasmussen argues that one main concern with the wrongness of discrimination, given that it is not always wrong, are its harmful outcomes. Some statements defended in part 1 of the book have a pervasive impact in this second part of the book. For example, in part one Lippert-Rasmussen states that discrimination is essentially comparative, and, as mentioned before, this completely determines the account of the wrongness of discrimination. To wit, according to this account, the wrongness of a discriminatory act is based on its effects, and not on any other intrinsic moral wrongness it may generate. In addition, a discriminatory act will be harmful if and only if the discriminatee is worse than she would have been had she not been discriminated. However, discrimination may be morally wrong for other reasons than the ones mentioned in Lippert-Rasmussen's approach in part 2 of the book. For instance, racist, sexist, male chauvinistic attitudes may be morally bad both for the discriminatee and for the discriminator. Or they may have no bad effects in the discriminatee, whilst remaining morally bad for the discriminator, in terms of attitudes, decisive reasons for action, and bias generally generated by false beliefs.

On this line of reasoning, discrimination based on inequalities may be morally wrong, not just because of the alleged injustice of inequalities, but also due to the fact that it emphasizes previous injustices, structural or otherwise. Lippert-Rasmussen is aware of that previous injustices aggravate the harm of discriminatory acts (55 and 62). However, the harm-based account defended by the author does not take into account moral wrongs other than foreseen harmful out-

comes to constitute the wrong-making property, (155). For example, discriminatory acts may generate unintended harms, and both these harmful byproducts, and the discriminatory act generating both types of outcomes, raise moral concerns. It seems to me that these aggravating factors are disregarded in Lippert-Rasmussen's account of the wrongness of discrimination.

If we consider it in more detail, we will see that in part one of the book Lippert-Rasmussen conceives indirect discrimination as a non-intentional mental state[1] (73). Accordingly, indirect discrimination may be wrong in light of its due outcomes. However, discrimination based on mental states may well be intentional, and therefore morally wrong not only in virtue of its outcomes, but of its reasons for action. Hence, if Lippert-Rasmussen agrees with the claim that indirect discrimination may well be equally harmful, we may add that this would not be solely due to its harmful outcomes, but also of its reasons for action.

Finally, in part 3 of the book, Lippert-Rasmussen introduces three so-called non-ideal themes: proportional representation in connection with punishment, discrimination on the labour market, discrimination in the private sphere, and, finally, racial profiling. He discusses them in light of his proposed account of discrimination, the desert-prioritarian account. The chapter on discrimination in the private sphere is particularly interesting.

Despite the set of issues that need clarification, and further development, *Born Free and Equal* is a worthwhile enjoyable read, and it sets a precedent for further and fruitful discussion on the somewhat neglected topic of discrimination in political philosophy.

Cristina Astier
Philosophy of Law Area
Department of Law
Pompeu Fabra University
Edifici Roger de Llúria, Ramon Trias
Fargas, 25-27 | 08005 Barcelona
cristina.astier01@estudiant.upf.edu

[1] The discussion on the wrongness of indirect discrimination remains open, and Lippert-Rasmussen comes back to it at the annex of chapter 6, at pages 177, and 178.

The Double Lives of Objects: An Essay in the Metaphysics of the Ordinary World, by Thomas Sattig. Oxford: Oxford University Press, 2015, 288 pages, ISBN 9780199683017 (hbk).

In *The Double Lives of Objects* Thomas Sattig defends an original and highly interesting account of ordinary objects like mountains, oaks, statues and people: *perspectival hylomorphism*. The account has a metaphysical part, *(quasi)-hylomorphism*, and a semantic part, *perspectivalism*. The author situates the account somewhere in between the two prevailing theories, *classical mereology* and *Aristotelian hylomorphism*, and argues that it is better placed than its contenders to preserve our common-sense conception of ordinary objects, offering a unified and compatibilist solution to a range of problems that challenge this view.

The structure of the book is clear: first, the basics of the theory are developed (chapters 1 and 2), and then the theory is extended and refined through its application to a series of issues that threaten our common-sense view of ordinary objects (chapters 3-8). Each chapter in this second part can be read independently of the others.

Let me outline Sattig's theory and stress some points I believe deserve special attention and further discussion.

Sattig presents his account as a fundamentally classical-mereological account with an Aristotelian *twist*. Like classical mereology, it understands *complex material objects* as mereological sums of smaller material objects but, against this view, it affirms that ordinary objects are not just material objects. On the other hand, like Aristotelian hylomorphism, it distinguishes between an ordinary object's *matter* and *form*, but it understands forms very differently.

Sattig's perspectival hylomorphism views *ordinary objects* as *compounds* of *material objects* and *K-paths*. Let us see what this means.

Sattig understands *material objects* in accordance with classical mereology, of which he presents several versions (depending on whether *temporal parts* are accepted or not) and claims that his framework can be developed using any of them. However, he mainly uses the three-dimensionalist version in which material objects cannot change their parts over time (this will be important). Accordingly, I will restrict myself here to this version. He also emphasizes that material objects have non-derivative spatiotemporal locations and

physical properties.

Now, let us see what *K-paths* are. We need to introduce several notions.

First, each kind *K* has associated a certain qualitative content, Φ^K, shared by all its instances (for example, for the kind *table* it mainly comprises functional properties).

Second, Φ^K is instantiated by material objects. Suppose that a material object *a* instantiates Φ^K, and suppose that *a*'s being ψ_1, *a*'s being ψ_2…and *a*'s being ψ_n jointly ground *a*'s being Φ^K. Then we say that this plurality of properties ψ_1, ψ_2... ψ_n completely realizes *K*.

Third, for any kind *K* there is a range of properties that can meaningfully be ascribed to *Ks*. They constitute its *sphere of discourse*.

Now we can characterize a *K-state* of a material object. For any kind *K*, a *K*-state of a material object is a complex, conjunctive, fact about the material object that obtains at a particular time. More precisely, a *K*-state (for some kind *K*) of a material object *a*, at a time *t*, contains two types of qualitative profile:

(1) The *K-meaningful intrinsic profile* of *a* at *t*. This contains:
The maximal conjunction of the facts that *a* exists at *t*, that *a* has α_1 at *t*, …, that *a* has α_n at *t*, such that (i) each α_i is an intrinsic qualitative property of *a*, and (ii) each α_i falls in the sphere of discourse of *K*.

(2) The *K-realization profile* of *a* at *t*. This is constituted by two types of fact.

(2.1) The maximal conjunction of the facts that *a* has ψ_1 at *t*, …, that *a* has ψ_n at *t*, such that properties ψ_1 …, ψ_n together completely realize *K* (i.e., the maximal conjunction of the facts about *a* that jointly ground *a*'s being Φ^K).

(2.2) The maximal conjunction of the facts that ψ_1 partly realizes *K*, …, that ψ_n partly realizes *K*.

(This last clause is crucial to the solution of the *grounding problem*.)

We can now introduce the notion of a *K-path*. Intuitively, whereas a *K*-state is the imprint (as Sattig says) of a kind *K* on a material object

at a particular time, a *K*-path is a series of imprints of *K* over time. Intuitively, a *K*-path is the life of a *K*.

More precisely, a *K*-path is a maximal series of *K*-states unified by *K-continuity*, *K-connectedness* and *lawful causal dependence*.

An important characteristic of *K*-paths is that they may have distinct material objects as subjects (remember that material objects do not change their parts over time). On the other hand, a material object may be a subject of distinct *K*-paths, even of distinct kinds.

Finally, Sattig states that an *ordinary object* is a transcategorial mereological sum of a material object and a *K*-path that has the material object as one subject (remember that a *K*-path can have more than one subject). Sattig calls them 'compounds'. Analogously to sums, the identity conditions of compounds just depend on the compounds' parts, irrespective of what these are and of how they are arranged.

Let me highlight a couple of consequences. First, this account yields a plenitudinous ontology. Just one example: consider a particular *Table*-path, *i*, and suppose that *i* has distinct material objects a_1, a_2, a_3 as subjects. Then, we have three different tables: the compound of a_1 and *i*, the compound of a_2 and *i*, and the compound of a_3 and *i*. Second, and this is a crucial aspect of Sattig's proposal, the qualitative profile of an ordinary object's material object (its *matter*) and the qualitative profile of the same object's *K*-path (its *form*) may diverge.

After presenting the metaphysical part of his account, Sattig compares it with its rivals. He views the discrepancy with regard to classical-mereological accounts as not being metaphysically substantive, just a metaphysical disagreement about the nature of some derivative objects. However, the discrepancy with Aristotelian accounts is, Sattig affirms, metaphysically substantive. For example, Aristotelian forms play an object-structuring and an object-generating role. This is not the case for *K*-paths.

Now, let me summarize Sattig's criticism of Aristotelian hylomorphism. He claims that the nature of its primitive *structuring composition operations* and their associated forms is mysterious: how can they be sensitive to particular, high-level kinds of objects and arrangements? For example, what explains the relevance to the application of a composition operation that five objects are such that

four of them are legs and the other a top and that they are arranged tablewise? In Sattig's opinion:

> Generating a new object is a metaphysically robust job. When a mechanism with this job is tuned to specific, high-level properties and relations, we expect an explanation of the mechanism in more basic terms [...] For how can something this fundamental be sensitive to something this derivative? (10)

I have some doubts about this criticism. Before explaining them, let me say that, for reasons of space, I can only present them briefly. A fuller development remains a task for another occasion.

My concern about Sattig's criticism is that his account seems to appeal to (in this case) a relation relevantly similar to Aristotelian composition operations: the relation of *subjecthood* between material objects and *K*-paths.

Suppose that the qualitative content of the sortal *table* states (I am simplifying) that tables have four legs and a top arranged tablewise.

Broadly speaking, according to the Aristotelian structuring composition operation associated with the sortal *table*, in order for a table to exist there have to be four legs and a top arranged tablewise.

Now, this seems to be relevantly similar to what happens in Sattig's framework. Broadly speaking, in order for a material object to be the subject of a *Table*-path it has to have proper material parts which are the subjects of four *Leg*-paths and one *Top*-path and it has to instantiate the tablewise arrangement (further conditions are required, but they are not directly relevant here).

It is true that in the case of Aristotelian accounts the successful application of the relevant structuring composition operation implies the existence of a table, and in the case of Sattig's account we still need to sum the material object and the *Table*-path to obtain a table. However, that the material object and the *Table*-path stand in the relation of subjecthood is a pre-requisite for this sum to result in the compound that is the table. Is this difference so decisive as to see Aristotelian composition operations as suspicious and mysterious, but not the relation of subjecthood? It would be interesting to know more about this relation in general and how it compares to Aristotelian composition operations.

After presenting q-hylomorphism Sattig introduces *perspectivalism*, a metaphysical semantics of the statements expressing our com-

mon-sense conception of objects. Sattig elaborates it in the form of a truth-theory stated in terms of q-hylomorphism.

First, he defends that we might adopt three different, unconnected, perspectives on ordinary objects: two common-sense perspectives, and the *absolute perspective* of fundamental metaphysics (which is not accessible from common sense). One of the perspectives of common sense is the *sortal-sensitive perspective* from which we represent ordinary objects in manners that are sensitive to the kinds to which they appertain. The other is the *sortal-abstract perspective* from which we represent ordinary objects in primarily spatiotemporal terms, irrespective of the kind to which they belong. From this perspective, for example, it is a platitude that (a) an object has a continuous spatiotemporal path, or that (b) there cannot be different objects at the same place at the same time, or that (c) an object cannot cease to exist in virtue of merely extrinsic causes. Sattig adds that this perspective is *fragmented and amorphous*, providing at most a partial principle of individuation. One of the examples Sattig uses to show this is the following: imagine a brick wall abstracting from all features making it a brick wall. Suppose one more brick is added. Does it merely receive an external attachment or does it increase its size? Sattig claims that spatiotemporal continuity is compatible with both options: *the object to which a merely external thing is added*, but also *the object which increases its size*, have a spatiotemporally continuous path.

Sattig offers the following reason for differentiating between the two common-sense perspectives. Psychological research indicates that infants represent objects in a primarily spatiotemporal way. However, adults seem to represent objects (also) as appertaining to sortals. Now, the most plausible explanation of this evolution is that, in fact, infants' object representation principles continue to be active in adults, and are the basis of common-sense platitudes like (a)-(c). After this, Sattig adds: given that these underlying principles are sortal-abstract (here he equates *sortal-abstract* with *spatiotemporal*, but this is the issue in question, as we will see), (a)-(c) should be seen as sortal-abstract, as well. This is a good reason, Sattig affirms, for differentiating between the two common-sense perspectives.

I have some doubts about Sattig's reasoning (as I said in the above case, I can only present them briefly here, and a fuller development remains a task for another occasion). The data from psychological

research he provides in the book (i.e., that infants mainly use spatiotemporal principles to individuate objects) also seem compatible with the thesis that there is just one human perspective on ordinary objects which is built up over the years: infants' spatiotemporal principles can be seen as the first step in the construction of a far more complex, but unique, sortal-sensitive, perspective. These principles would then also be part of the sortal perspective of adult human beings.

Why should we prefer Sattig's proposal to one that accepts a unique perspective which develops step by step over the years?

Sattig emphasizes at several places that these principles seem to apply to all ordinary objects independently of the specific properties that make them chests of drawers, roses, mountains or dogs. They would be, then, general sortal-abstract principles. But this does not seem to me to be as clear as he claims. Intuitively, a tree, a person or a table is a tree, a person or a table because (apart from other requirements) it obeys principles of the sort of (a)-(c). Intuitively, I would say that a table is a table, in part, because, for example, it cannot jump between distant places from one moment to the next and it cannot cease to exist for purely extrinsic causes. Moreover, that these principles apply to all ordinary objects might just mean that they are common to all sorts.

Now, Sattig's next step is to defend that to a type of perspective there corresponds a *mode of predication*. By adopting the sortal-sensitive perspective, we employ the *formal* mode of predication. By adopting the sortal-abstract perspective, we employ the *material* mode of predication. By adopting the absolute perspective, metaphysicians employ the *absolute* mode of predication. Formal descriptions track properties contained in an ordinary object's *K*-path, whereas material descriptions track properties instantiated by an ordinary object's material object. For example, when considering a table's formal persistence (from the sortal-sensitive perspective) we track the temporal trajectory included in its *Table*-path; however, when we consider the material persistence (from the sortal-abstract perspective) of the same table we track the temporal trajectory of its material object.

Sattig emphasizes that the key feature of perspectival hylomorphism is that it allows *perspectival divergence* based on *hylomorphic di-*

vergence (ordinary objects live double lives!): an ordinary object may have different profiles from different perspectives because the profile of its material object and the profile of its *K*-path may take different directions. For example: suppose that material object a_1 exists at t_1 but not at t_2 and that material object a_2 exists at t_2. Moreover, suppose that a *Table*-path *i* includes the fact that a_1 exists at t_1 and that a_2 exists at t_2. Then, among others, there is table *o*, the compound of a_1 and *i*. Now, when we say, from the sortal-sensitive perspective, using the formal mode of predication, that '*o* exists at t_2' we are saying something true, and when we say, from the sortal-abstract perspective, using the material mode of predication, that '*o* does not exist at t_2' we are also saying something true.

I have some doubts related to the two following theses that Sattig proposes: the thesis that the sortal-abstract perspective is, in Sattig's words, fragmented and amorphous and the thesis that the mode of predication associated with this sortal-abstract perspective, the material mode of predication, tracks the properties of ordinary objects' material components, i.e., of material objects. As in the above cases I can only present my doubts in outline here: it is not clear to me how much of this sortal-abstract perspective of common sense Sattig wants to vindicate. From what he says in the book the answer seems to be "as much as possible". However, given the two theses mentioned, this does not seem an easy task. Let me just present one reason: on the one hand, our material predications (made from the sortal-abstract perspective) about the persistence of an object through time will show that our sortal-abstract perspective is fragmented and does not include any determinate, precise, persistence conditions of objects. On the other hand, the persistence conditions of material objects, in terms of which these sentences will be evaluated as true or false, are determinate, as they are the persistence conditions of mereological sums. In fact, this tension can be exemplified using the cases Sattig presents to illustrate the indeterminacy of the sortal-abstract perspective. I will use the one I have reproduced above: the example of the brick wall to which one further brick is attached. From the sortal-abstract perspective we would describe the case as one in which it is indeterminate whether the brick wall has something externally attached to it or is increasing in size. However, the sentences we would use in the description would be evaluated

in terms of what happens to the material object that is the material component of the brick wall. As material objects cannot change their parts, this will determine that the brick wall does not change in size.

In the remaining chapters Sattig defends his theory, arguing that perspectival hylomorphism offers the best solution to a series of problems that threaten our conception of ordinary objects. I do not have space here to discuss his specific solutions to every specific problem. However, I would like at least to point out one recurring worry I have with Sattig's characterization throughout the chapters of the sortal-sensitive perspective of common sense. I doubt that some of the theses that he claims to be in accordance with such a perspective are really so: for example, the claim that two objects of the same sort can coincide.

In chapters 3 and 4 Sattig discusses *paradoxes of coincidence*, *cases of fission* and *cases of intermittent existence*. He argues that the theses seemingly leading to paradoxical results express, in fact, different perspectives (some the sortal-sensitive perspective, some the sortal-abstract perspective) and therefore, contrary to first appearances, they are compatible.

In chapter 5 the framework is refined and applied to modal issues. In a nutshell, material objects exist in different possible worlds whereas *K*-paths are worldbound, having *counterparts* in other possible worlds. *Ordinary objects* are compounds of transworld material objects and worldbound *K*-paths. Moreover, formal *de re* modal attributions are understood in terms of counterparts of the objects' *K*-paths, and material *de re* modal attributions in terms of the objects' material components.

In chapter 6 Sattig states that friends of coincidence have to accept that the actual world is *indeterministic* on *a priori*, mundane grounds; and this is absurd. Sattig's solution: questions of determinism concern just qualitative properties of material objects.

Chapter 7 offers an account of certain *indeterminate* properties of objects. Sattig introduces *multiple superimposed individual forms* and analyses indeterminacy as formal indeterminacy.

In the last chapter Sattig gives an account of certain puzzling relativistic properties of ordinary objects appealing to different, compatible, perspectives we may take on these objects.

Let me finish by saying that I believe Sattig does an excellent job

in the search for a much wanted theory that combines the virtues of opposing theories. I cannot recommend this book highly enough.

Marta Campdelacreu
Universitat de Barcelona, LOGOS
marta_campdelacreu@ub.edu