Contents

Organon F 27 (3) 2020: 282–301
https://doi.org/10.31577/orgf.2020.27301

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

What We Know about Numbers and Propositions and How We Know It

Scott Soames*

Received: 30 December 2019 / Accepted: 24 January 2020

Abstract: The paper sketches and defends two instances of the strategy *Let N's be whatever they have to be to explain our knowledge of them*—one in which N's are natural numbers and one in which N's are propositions. The former, which makes heavy use of Hume's principle and plural quantification, grounds our initial knowledge of number in (a) our identification of objects as falling under various types, (b) our ability to count (i.e. to pair memorized numerals with individuated objects of one's attention), (c) our (initially perceptual) recognition of plural properties (e.g. being three in number), and (d) our predication of those properties of pluralities that possess them (even though no individuals in the pluralities do). Given this foundation, one can use Fregean techniques to non-paradoxically generate more extensive arithmetical knowledge. The second instance of my metaphysics-in-the-service-of-epistemology identifies propositions (i.e. semantic contents of some sentences, objects of the attitudes, and bearers of truth, falsity, necessity, contingency, and apriority) with certain kinds of purely representational cognitive acts, operations, or states. In addition to providing natural solutions to traditionally unaddressed epistemic problems involving linguistic cognition and language use, I argue that this metaphysical conception of propositions expands the solution spaces of many of the most recalcitrant and

* University of Southern California
University of Southern California School of Philosophy, 3709 Trousdale Parkway, MHP 113 (223 Stonier Hall), Los Angeles, CA 90089, U.S.A.
soames@usc.edu

long-standing problems in natural-language semantics and the philosophy of language.

Keywords: Arithmetic; cognition; knowledge; natural numbers; plural properties; plural quantification; propositional attitudes; propositions; representation; semantics.

All of us know a lot about propositions and numbers. We know that *belief*, *assertion*, and *knowledge* relate agents to *what they believe*, *assert*, *or know*—i.e. to propositions. Since what is believed or asserted can be true or false, we know that propositions are bearers of truth value, some are necessarily true, or false, while others are only contingently so. We also know they are asserted by uttering sentences. We know that natural numbers are the subject matter of arithmetic, which we learn as children. Our knowledge of propositions and numbers is commonplace. But it is also mysterious. We are philosophically in the dark about what they are. If asked *Which entity is the number 3? or the proposition that the sun is a star*? centuries of philosophical investigation haven't helped much. We ought to be able to do better.

I begin with a puzzlement our knowledge of numbers generates that isn't generated by our perceptual knowledge of trees. Aside from attempts to explain why there is no basis for radical skepticism about knowledge of the external world, philosophers haven't been overly perplexed by the fact that we know there are trees. But unlike trees, which we can see, philosophers generally agree that we can't see natural numbers. Although I know of one philosopher who has suggested that we can see certain sets, *the members of which can be seen*, I am not aware of any who have held that we can see numbers.[1] Even if we can see some sets, the usual set-theoretic conceptions of natural numbers don't encourage the idea that we can perceive them.

[1] After usefully reviewing the psychological literature on (physical) object perception, Maddy (1980) argues (i) that some perceptual beliefs, e.g. *that there are 3 eggs left in the carton*, are numerical in nature and (ii) that in her example the set containing the eggs is a constituent of the belief (which is judged to be 3-membered). Although I agree with (i), I will argue that our knowledge of natural numbers is better explained in another way.

Frege's idea, that we can find out what numbers are by finding out what best explains our knowledge of them, is compelling. In sections 58 and 60 of *The Foundations of Arithmetic*, he says that nothing we can picture or imagine seems to be an apt candidate for being the number 4. But he isn't deterred. Although we can, in his words, *form no idea of the content of a number term*, he insists that this no reason for denying that it has a content (see section 58 of Frege [1884] 1950). Rather than considering the term in isolation, we should, he thinks, ask what it contributes to meaningful sentences containing it. We must "*Always* [he says] *keep before our eyes a complete proposition [*sentence*]. Only in a proposition* [sentence] *have the words really a meaning*" (Frege [1884] 1950, section 60, p. 71). This is his "context principle," to which we can add the related principle: *Natural numbers are whatever they have to be in order to explain our knowledge of them.* There are many sentences containing numerical terms that express propositions we know. The strategy for finding out what natural numbers are is to investigate which assignments of meanings and referents to these terms best advance our ability to explain our knowledge of the truth of the propositions expressed by numerical sentences.

By 'our knowledge', I mean everyone's knowledge—children who know only a little, adults who know more, and number theorists who know much more. I presume this vast population shares a fair bit of common knowledge, even though some know more than others. Since no one knows all arithmetical truths—or even any collection from which all others could be validly derived—the set of arithmetical truths will outstrip all actual arithmetical knowledge. Still, we should be able to explain possible extensions of the knowledge we now have. I most want to know (i) how we achieve *any* knowledge of numbers at all, and (ii) how, with instruction, we can acquire more. In short, I am looking for a realistic starting point for a plausible account of our arithmetical knowledge.

One part of that starting point is what Frege called "Hume's Principle." It says that *the number of X's* is the same as *the number of Y's* iff the X's and Y's can be exhaustively paired off without remainder. E.g., *the number of universities* at which I have been a regular faculty member—Yale, Princeton, and USC—can be exhaustively paired off (without remainder) with the fingers I am now holding up. So, the number of *universities* at

which I have served is the same the number as the number of *fingers* I am holding up. Both are three (in number). What is this property, *being three (in number)*, predicated of? It's not predicated of any of my past faculty homes; neither Yale, Princeton, nor USC, is three (in number). It is not predicated of the set that contains just them; since the set is a single thing, it's not three either. Like the property *being scattered*, the property *being three* is plural. It is a property that applies, not to a *single* instance of any type of thing, but to *multiple things of a given type* considered together. My former Ph.D. students are scattered around the world, even though no one of them *is* scattered around the world, and the set containing them isn't scattered either.

With this in mind, consider the hypothesis that each natural number N greater than or equal to 2 is the plural property *being N (in number)*, and that the number 1 is a property applying to each individual thing considered on its own.[2] Zero is a property that doesn't apply to anything, or things. Natural numbers are such *cardinality properties*. In the beginning, we gain knowledge of them by counting. Imagine a child inferring that I am holding up three fingers from her perceptual knowledge that x, y, and z are different fingers. Having learned to count—by memorizing a sequence of verbal numerals—she concludes that the fingers are three in number. She pairs off, without remainder, the fingers I am holding up with the words she speaks—'one', 'two', and 'three'—thus ensuring that the fingers and the numerals "have the same number" in Frege's sense. The number they share is designated by the numeral, "three," that ends the count; it is the property *being three* which applies to these fingers.

This is the germ of an idea that combines the best of the Frege-Russell reductions with a striking, but incompletely developed insight in section 1 of Wittgenstein's *Philosophical Investigations*. The book begins with a quotation from Augustine.

> When they (my elders) named some object, and accordingly moved toward something, I saw this and I grasped that the thing

[2] This way of thinking of natural numbers grows out of two path-breaking works, one—(Boolos 1984)—by my former teacher, and the other—chapter 4 of (Gomez Torrente 2019)—by my former student.

> was called by the sound they uttered when they meant to point it out. Their intention was shown by their bodily movements, as it were the natural language of all peoples: the expression of the face, the play of the eyes, the movement of other parts of the body, and the tone of voice, which expresses our state of mind in seeking, having, rejecting, or avoiding something. Thus, as I heard the words repeatedly used in their proper places in various sentences, I gradually learnt to understand what objects they signified. (Wittgenstein 1958, section 1)

Wittgenstein uses the passage to illustrate a general conception of language he rejects—a conception in which naming is the essence of meaning. One reason he rejects this conception involves an imagined priority in introducing words into a language, and in learning a language. *First* comes our awareness of things in the world, *then* comes our use of words to talk about them. In learning the word, we first focus on candidates for its referent. Then we converge on the single candidate that best makes sense of the sentences they use containing the name.

Having sketched the general picture he wants to reject, Wittgenstein immediately jumps to a use of language which, he thinks, doesn't conform to it. He says

> Now think of the following use of language: I send someone shopping. I give him a slip marked "five red apples." He takes the slip to the shopkeeper, who opens the drawer marked "apples"; then he looks up the word "red" in a table and finds the color sample opposite it; then he says the series of cardinal numbers—I assume that he knows them by heart—up to the word "five" and for each number he takes an apple of the same color as the sample out of the drawer. It is in this and similar ways that one operates with words […] "But how does he know…what he is to do with the word 'five'?" […] [W]hat is the meaning of the word "five"? —No such thing was in question here, only how the word "five" is used. (Wittgenstein 1958, section 1)

This emphasis on the use of the *numeral* 'five', rather than its referent, is illuminating. But the proper lesson *isn't* that its meaning is its use; the

meaning of the word 'five', which is also its referent, is the *property being five*, which, being true of Wittgenstein's apples, isn't a use of anything. The lesson is that our use of the numeral in counting makes us aware of the property, which becomes cognitively associated with the numeral. *First the use, leading to awareness of something to be named; not first the awareness of number, and then the decision to name it.*

Counting, emphasized in the passage, establishes an epistemic foothold on a vast domain that none of us will ever actually count. Most of us know how to count to a trillion. But some of us don't know a verbal numeral for the number that comes after *nine hundred ninety nine trillion, nine hundred and ninety nine billion, nine hundred and ninety nine million, nine hundred and ninety thousand, nine hundred and ninety nine.* Still, most people have mastered the system of Arabic numerals, in which each natural number has a name, even though no one will ever use them all. These names are rigid designators the referents of which are fixed by descriptions that are implicitly mastered by those who understand them.

They can each be taken as designating a distinct cardinality property, as long as we don't run out of multiples to bear those properties. This might seem problematic, since it is likely that there are only finitely many electrons in the universe, and so only finitely many multiples of concrete things. But this needn't be problematic because we aren't restricted to counting concrete things. We can also count multiples which include plural properties among them, including numbers we have already encountered. This ensures that there is no end to larger and larger multiples, and so no end to our cardinality properties.

This picture gives us a way to explain our knowledge of numbers. Consider the child inferring from her perceptual knowledge that the number of fingers I am holding up is 3. At first, she does this by counting—saying the first three positive numerals—pairing them off with the fingers I am holding up. In time, counting won't always be necessary, because she will recognize at a glance when she is perceiving trios of familiar types. At that point she has the concept, *being a trio of things* (of some type), which is the plural property that is the number 3. She learns a few other small numbers initially by counting, but eventually by perceptual recognition, and forming perceptual beliefs. She can perceptually recognize instances of these numbers, even

though counting will remain her fallback when in doubt, or when the multiples increase in size. In this way, much of our knowledge of natural numbers is knowledge of plural properties grounded initially in perception, in cognitive recognition of things being of various types, and in cognitive action—counting the items falling under a given concept by reciting the relevant numerals while focusing one's attention on different individuals of the given type.

One doesn't *first* learn what numbers are, and *then* use them to count. One first learns to articulate numerals while pairing them with sequences of things. One begins to recognize and refer to numbers when one has mastered this practice and integrated it into one's cognitive life. In saying, "The number of blue books on the table is four, but the number of red ones is only three," one uses the numerals to attribute cardinality properties of multiples. The properties one attributes are numbers, which exist independently of us and of our language, but which we come to cognize only in virtue of the linguistic, and other symbolic, routines we have mastered.

Now for some refinements. I take the word 'three' to be analogous to the word 'blue'. Both can perform three grammatical functions: First, they can be used to designate properties of which other properties are predicated, as in and *Blue is the color of a cloudless sky at noon* and *Three is the number of singers in the group.* Second, they can be used to form predicates, as in *The sky is blue* and *We are three*, said by Peter, Paul, and Mary in answer to the question *How many are you?* Third, they can modify predicates, as in *There is a blue shirt in the closet* and *There are three singers on stage.* The numeral 'three' designates a plural property that applies to Peter, Paul, and Mary without applying to any one of them; the compound property *being three singers on the stage* applies to some individuals who are (collectively) three iff each is a singer on the stage.[3]

In a formal language, these distinct uses of the word 'three' might be regimented into uses of different words, but that needn't concern us here. The idea that properties, which we predicate not only of individuals but

[3] To say that there are at least 3 singers on the stage is to say that some individuals who are three in number are each on the stage. To say that there are exactly 3 singers on the stage is to say that some individuals who are 3 in number are each singers on the stage, but no singers on the stage are collectively greater than 3.

also of properties opens the door to paradox, unless restrictions are adopted. But the danger is general, and not, I think, specific to plural cardinality properties. There is, of course, no property that is true of any property p iff p isn't true of itself. But there is no need for one. I.e., there is no need for every meaningful predicate, including "*is a property that is true of each property that is untrue of itself and of no others*," to express a property. When it comes to the plural cardinality properties needed as natural numbers, no property, except the degenerate case of the number 1, is true of itself, because no natural number other than 1 is true of any single thing. There are, of course, plural cardinality properties that are true of some Fs *of which they are not one* as well, we may suppose, as plural cardinality properties that are true of some F's *of which they are one*. But this is also unparadoxical. Indeed, it would seem that all plural cardinality properties greater than or equal to 2 fall in both classes. Thus, it should be possible to explain our knowledge of arithmetic by taking natural numbers to be plural cardinality properties without paradox. Of course, no derivation, from logic itself, is contemplated.

What about a version of Frege's worry in section 46 of the *Foundations of Arithmetic*? While looking at Peter, Paul, and Mary standing on the stage next to The Beatles, I may say, equally, the number of singing groups on the stage is 2 or the number of singers on the stage is 7. Indeed, I see 7 singers and I see 2 singing groups. I am *not* saying that any things I see are both 2 and 7 in number. One might think this if one thought that the 7 singers were identical with the two groups, but they aren't. Although the 7 singers *constitute* the 2 singing groups there is no genuine identity here. For one thing, the singers were all much older than the groups, while the groups weren't much older than the groups; indeed neither group was much older than the other. Whenever we count, the items counted must already be individuated. The number 3 is the plural property applying to all and only those individuals x, y, and z none of which is identical with any other. In saying this, we presuppose the individuation we need when stipulating that the values of the variables. Nothing more is needed to predicate the plural property. Given this, we can accommodate Frege's observation in section 46.

So far I have talked only about early stages of our knowledge of natural numbers, some of which is perceptual belief that qualifies as knowledge. If

one's knowledge that x and y are fingers is perceptual, and one's knowledge that x isn't y is too, then one's knowledge that x and y are two things is also perceptual. If the fingers had been painted *blue* one could truly say, not only that *one sees that those fingers are blue*, but also that *one sees that they are two in number*. If two people are standing at a distance from someone holding up two blue fingers, one person, who has trouble making out what is displayed, might ask *Do you see the color of those fingers*? or *Do you see the number of those fingers*? The one with better vision might reply, *Yes, I see their color; they are blue* or *Yes, I see the number of those fingers; they are two*. So, there is a more or less ordinary sense of 'see' in which we can truly say that some color properties and some natural numbers, i.e. plural cardinality properties, can be seen. Don't wring your hands over this. If philosophy is worth doing, it should sometimes provide surprising, even shocking, knowledge. Here, it is knowledge about some of our knowledge of numbers.

Systematic arithmetical knowledge—e.g. of axioms and logical consequences of Peano Arithmetic—is more complicated. It isn't all *logical* knowledge of the sort Frege imagined. If natural numbers are *cardinality properties*, logic *alone* can't guarantee that there are any individuals, multiples, or distinct cardinality properties of multiples, let alone infinitely many. But we can use logic plus *updated versions of Frege's definitions* involving plural properties rather than sets to extend our knowledge of natural numbers. The definition of *successor* says that the plural property *N is the successor of the plural property M iff some Fs, of which a given object o is one, are N in number, while the Fs, excluding o, are M in number*. Given definitions of *zero* and *successor* we can define *natural numbers* as *plural properties of which every property true of zero and of the successor of anything it is true of, is true.*[4]

From this plus our initial perceptually-based knowledge, we can derive arithmetical truths. We can come to know that zero isn't the successor of anything by observing that if it were, some property true of nothing would be true of something. We learn that no natural number M has two successors by observing that otherwise there would be plural properties N1 and N2 such that the N1s *can't* be exhaustively paired off with the N2s, even

[4] See pp. 435–36 of Boolos (1984) for further discussion.

though there are objects o_{N1} and o_{N2} such that the *N1s excluding* o_{N1} and the *N2s excluding* o_{N2} are both M in number, and so can be paired off. Knowledge of the companion axiom, *that different natural numbers N1 and N2 can't have the same successor*, is explained in the same way. The axiom that every natural number has a successor is seen to be true when we realize that the plural cardinality properties we arrive at by counting can themselves be included in later multiples we count. This ensures we can always add one to any Fs of which a plural cardinality property M we have already reached is true. In this way, plural cardinality properties can help us explain not only our earliest knowledge of natural numbers, but also how systematic knowledge of elementary number theory can be acquired.

There are, of course, other ways of expanding the meager knowledge of arithmetic acquired in kindergarten. Most of us learned addition, subtraction, multiplication, and division without being exposed to Peano's axioms. The efficient, user-friendly routines we mastered are compatible with the perspective advocated here.

Is this perspective really correct? At least it is more promising than set-theoretic accounts of natural numbers. Paul Benacerraf's original problem alerted us the fact that any reduction purporting to tell us what natural numbers really are—as opposed to what, for some purpose, it is convenient to take them to be—must provide a good reason for selecting one set theoretic-reduction from among the many different ways of identifying natural numbers with sets (Benacerraf 1965). Although each identifies individual numbers with sets that differ from those provided by other reductions, the different reductive systems do an equally good job of preserving all arithmetical truths. So, if that were the only criterion for justifying a reduction, we would have no reason for thinking that any of those reductions is uniquely correct. We might even have reason for doubting that *any* is correct. Surely, one may think, if the number three is *identical* with some set, there should be a reason it is one in particular, rather than any others. But, Benacerraf plausibly suggests, no set-theoretic reduction provides such a reason.

If one believes, as I do, that there really are natural numbers, which all of us know about, then we should look for the best explanation of our knowledge. I suspect that Benacerraf was right in suggesting that set-theoretical reductions won't provide it. We do, of course, have knowledge

of sets. But it doesn't, I think, come directly from anything as immediate as counting and perception. Rather, I suspect, it arises from activities in which teams, committees, and groups of coordinated individuals collectively succeed in doing things than no individual does—like winning a football game or carrying piano too heavy for any one of them. After we have admitted these, I suspect we come to recognize collections of things that are noticeably similar in some way, even though they may not do anything. Later, it occurs to us that we have no reason to exclude arbitrary collections of things. At this point the axioms of set theory can be considered and accepted.

Even then it is hard to take seriously the idea that natural numbers are sets—partly for Benacerraf's reasons, but also because by the time children and young adults have reached the level required to appreciate set-theoretic abstraction, the natural numbers have already been cemented in their minds as plural properties. It is tempting to put it this way: just as the color blue is naturally understood to be the property commonly possessed by this, that, and the other individual blue thing, so the number three is naturally understood as the property commonly possessed by this, that, and the other trio, each of which is three in number. But one must be careful. This way of expressing the idea can be misleading. It makes it sound as if pairs, trios, and multiples in general were a kind of thing. They aren't—at least in the sense that plural cardinality properties are true of pairs, trios, and multiples in general. When we say that Peter, Paul, and Mary are three, *nothing of any kind is said to be three. Being three* is a property that is never correctly predicated of anything. It is a property correctly predicated of some *things*—e.g. Peter, Paul, and Mary—without being predicated of them individually. Of course natural numbers are not all the numbers there are. What about all the others? Many at least seem to be constructions out of natural numbers. The questions, *What sort of constructions?* and *How are they known?* might be vitally important. But at least we have a start.

1. Propositions

I now turn to propositions, the dominant conception of which identifies them with sets of possible world-states at which sentences are true. What

is it for a *sentence S* to be true at w? Well, what is it for me to be a Slovak at w? It is for w be a state, which, if the world were in it, I would be a Slovak. So, one might think, for S to be true at w is for w to be a state, which if the world were in it, *S would be true*. But that won't do, because the truth value of S *at world-states at which S means nothing, or something other than what it actually means, is irrelevant to its actual truth conditions*. What we ought to say when doing semantics is that for S to be true at w is for w to be a state which, if the world were in it, *the proposition we actually use S to express would be true*. But saying that requires knowing about propositions, their relationship to sentences, and what it is for *them* to be true at world-states, *before* we give the truth-conditions of *sentences* at world-states.

One could simply *stipulate* that propositions are sets of world-states, and that for one to be true at w is for w to be a member of it. But that creates more problems than it solves. First, it entails that necessarily equivalent propositions are identical, which misrepresents propositional attitudes. Second, it denies the plausible idea that propositions have truth values *because* they represent things as being certain ways—e.g. red or round—and so are true when the things are red or round. By contrast, the question *What does the set of worlds* $w_1 ... w_n$ *represent*? is nonsensical. It doesn't represent anything. Third, the identification of propositions with sets of world-states provides no explanation of what it is to entertain propositions, to believe them, or to describe others as believing them. Fourth, it inverts conceptual priorities. Instead of using propositions to define world-states, it treats world-states as unexplained primitives and uses them to define propositions. Finally, it misrepresents ordinary cognitions of all cognitive agents, no matter how primitive, as being about world-states, rather than about what we, and they, see, hear, taste, touch, and cognitively interact with in simple ways. Thus, what passes for knowledge of propositions, on the dominant conception, isn't knowledge. At best sets of world-states can, for some limited purposes, model propositions. But they aren't the real things.

Nor are what Frege and Russell called 'propositions' (see chapter 2 of Soames 2010). For Russell, propositions were mysteriously "unified" combinations of objects, properties, and propositional functions that are said to be true iff the properties are true of the relevant objects, or propositional

functions. For Frege, they are mysteriously "unified" combinations of "complete" or "incomplete" senses said to be true iff the concepts presented by incomplete senses are true of (lower-level) incomplete senses, or (at the lowest level) of objects. Though each account suffers from its own difficulties, they share a crippling problem. No set, sequence, or formal structure, of, in Russell's case, objects and properties, or, in Frege's case, complete or incomplete senses, represents, on its own, anything as being any way. Not being representational, Frege-Russell propositions can't be the source of intentionality. Nor do they have truth conditions in virtue of any natural relation they bear to us. Theorists who use them treat them as models, which they *interpret* as being true in specified conditions. Unlike these models, of which ordinary agents know nothing, real propositions *are* the interpretations agents assign to utterances (Soames 2010, chapter 5).

This leads to a second defect. Traditional conceptions of propositions don't tell us what it is to entertain or believe them, or how agents acquire knowledge of them. This is crucial because propositions impose conditions on minds that entertain them that are more fine-grained than the truth conditions they impose on the world. To miss this is to miss their *epistemic essence*. Traditional conceptions of propositions also miss the *semantic essence* of sentences that express them. Just as proponents of traditional conceptions fail to explain what they call *grasping a proposition* amounts to, so they fail to explain what it means for sentences to semantically *express* propositions (Soames 2016).

Finally, no traditional conception of propositions fully accommodates their hyper-intensionality. While the deficiencies of the possible-worlds conception are legendary, the puzzles posed for Russellian and Fregean propositions by current analyses of names, natural kind terms, indexicals, and pronouns functioning as variables are well-known. Despite decades of effort, only limited progress has been made on the classic puzzles of Frege, Mates, Kripke, Perry, and Church (Frege [1893] 1952; Mates 1952; Kripke [1979] 1988; Perry 1977, [1979] 1988; Church [1954] 1988). Although linguistic science has made great progress in the last 70 years, the semantics and pragmatics of hyperintensional constructions aren't among its triumphs. They aren't because we lack a clear and widely accepted idea of what propositions are. Our empirical shortcomings with hyperintensionality are due to our

metaphysical and epistemological cluelessness. To make empirical progress, we must attack foundational issues.

We start from the principle that agents are the source of intentionality. Agents represent things as being various ways when they perceive, visualize, imagine, or otherwise think of them as being those ways. Propositions are repeatable, purely representational cognitive act or operation types. When one perceives or thinks of B as hot, one predicates *being hot* of B, thereby representing it as hot. The act represents B as hot in a sense similar to the derivative senses in which acts can be irresponsible. Roughly, an act is irresponsible when to perform it is to neglect one's responsibilities. A similar derivative sense of *representing* allows us to assess cognitions. When to perceive or think of o as P is to represent o as it really is, we identify an entity, a particular cognition plus a property it has when it is accurate. The entity is a proposition, which is the repeatable mental act type of representing o as P. The property is truth, which the act has iff to perform it is to represent o as o, in fact, is.

Entertaining, i.e. performing, is the attitude on which other attitudes are conceptually based. To *judge* that B is hot is to perform the predication in an affirmative manner, forming or strengthening one's disposition to act, cognitively and behaviorally, toward B in ways conditioned by one's experience with hot things. To *believe* that B is hot is to be disposed to judge it to be. To *know* that B is hot is for B to be hot, to believe that it is, and to be safe in so believing. Since believing a proposition p doesn't require cognizing p, *any organism that can perceive or think of the objects and properties in terms of which p is defined can believe p*—whether or not it can predicate properties of p or think about p at all. Knowing things about propositions involves distinguishing one's cognitive acts from one another. Self-conscious agents who can do this can ascribe attitudes to themselves and others, and predicate properties of propositions. Focusing on their own cognitions, they identify distinct propositions as different thoughts, which leads them to conceive of truth as a form of accuracy. How a proposition represents things is read off the acts with which it is identified, from which we derive its truth conditions. P is true at world-state w iff were w actual, things would be as p represents them—where what p represents is what any conceivable agent who entertains p would represent. Since this doesn't vary

from world-state to world-state, p's truth conditions are essential to it. No one has to entertain p for p to be true (see chapter 2 of Soames 2015).

This view explains how an organism without the ability to think about propositions can know or believe them. It also explains how sophisticated agents acquire the concept *proposition*, and come to know things about them by monitoring their own cognitions. It even gives the beginning of an account of what it is for a proposition p to be the meaning of a sentence S, as well as what it is for speakers to at least minimally understand S. Roughly, it is for speakers to use S to perform p. Learning a language involves learning how to use its sentences to perform the same propositions that others do. One who is competent with the sentence 'Kripke is human' uses the name to refer to the man, the noun to refer to humanity, and the phrase 'is human' to predicate the property of the man—thereby performing the proposition that is the semantic content of the sentence. That's not the only proposition one thereby entertains. *Using the sentence to predicate humanity of Kripke* is itself a purely representational act, and so counts as a proposition p*. Since to perform p* is to perform p, but not conversely (just as to drive to work is to travel to work, but not conversely) the two propositions are cognitively distinct, despite representing the same thing in the same way.

The reality of *representationally identical but cognitively distinct propositions* resolves many worries about hyperintensionality (Soames 2015, 39–45 and chapters 3–5). Consider uses of the 'Hesperus' and 'Phosphorus', which are unusual among names in imposing rich conditions on what it takes to understand them. Those who use them are expected to know that uses 'Hesperus' presuppose that it stands for something visible in the evening, while uses of 'Phosphorus' presuppose that it stands for something visible in the morning. To mix this up is to misunderstand them.

Suppose A utters "Hesperus is Phosphorus," addressing B, when both are presupposed to understand the names. A's utterance simultaneously asserts 2 *representationally identical propositions*, one which merely predicates identity of Venus and Venus, placing no restrictions on how the predication targets are cognized, the other which is entertained only by those who identify Venus using the two names. The former proposition, which is the semantic content of the sentence uttered, is necessary and knowable

apriori. The latter proposition is necessary but knowable only aposteriori (since knowing it to be true requires knowing the names to be coreferential). Although both propositions represent Venus as being Venus (and only that) the hearer B can draw further conclusions from A's utterance. Knowing that A presupposes that both of them understand the names, B knows that A realizes he will be taken to be committed to the claim that *the object Hesperus, visible in the evening sky, is identical with the object Phosphorus, visible in the morning sky.* Since A anticipates this, B correctly concludes that A *asserts* the descriptively enriched proposition, along with the uninformative, bare proposition.

The fact that the descriptively enriched proposition is contingent doesn't prevent A from correctly saying "Necessarily Hesperus is Phosphorus." The proposition embedded under the modal operator *isn't* descriptively enriched, because taking the names to designate referents *actually* seen at certain times (which is what understanding the names insures) provides no information about when the referents are seen at *possible world-states.* This explains why uses of names often contribute descriptive assertive content to clauses under 'assert', 'believe' or other attitude verbs, without making such contributions under modal operators (Soames 2015, 84–88).

Next consider natural kind terms. 'Water' and 'heat' are directly referential designators of kinds—one involving hydrogen and oxygen, one involving the motion of molecules. In each case, a kind K is the semantic content that a term G contributes to propositions semantically expressed by sentences containing G. Given this plus the idea that the semantic content of G is G's meaning, one is tempted to think (i) that K is the meaning of G, (ii) that knowing this is knowing *what G means*, and (iii) that since knowing *what G means* is the same as understanding G, understanding G is knowing that it means K. This conclusion is false (Soames 2015, 88–89).

Understanding requires more than minimal competence with the term, which is simply the ability to use it with its semantic content. To understand a term is to have the knowledge and recognitional ability to use it to communicate in ways widely presupposed in the linguistic community. This dynamic, illustrated by 'Hesperus' and 'Phosphorus', but rarely found with ordinary names, is nearly always present with natural kind terms. Understanding them—in the sense needed to use them to communicate in ways

widely presupposed by members of one's linguistic community—requires, in the case of 'water', knowing that users presuppose that it stands for something that can take the form of a colorless drinkable liquid that falls from the sky in rain and that is necessary for life. Similar observations apply to the terms 'heat', 'light', and 'red', the (full) understanding of which may sometimes require ability to recognize instances of the kinds.

Understanding in this sense is *not* a semantic notion in the sense of theories of semantic content. Our ordinary notions of *understanding an expression E* and *knowing what E means* track information commonly presupposed by most who use E. For a semantic theory that assigns a semantic content K to E to be correct, most minimally competent speakers must use E with that content, which must typically appear in the contents of speech acts performed using E. Widely shared presuppositions, which often carry extensive non-semantic content, distribute that content in the contents of speech acts involving the relevant expressions according to general pragmatic principles.

We can illustrate this with 'water' and 'H_2O'. The proposition *that water is water* predicates identity of the kind water and itself, and so is knowable apriori. The proposition *that water is the substance molecules of which are made up of one hydrogen atom and two oxygen atoms* is both distinct from that proposition and non-trivial because it involves the content of a definite description. The proposition semantically expressed by the sentence 'Water is H_2O' depends on whether 'H_2O' is a name or an abbreviated definite description. Suppose it is a Millian name the understanding of which requires associating it with some conventional information. Suppose further that common users—most of whom aren't educated in chemistry—must, to be counted as understanding it, know that it is widely taken to stand for some kind of chemical compound involving hydrogen and oxygen. (Nothing more detailed than that.) Then, the sentence *semantically* expresses the same trivially true proposition that 'Water is water' does. But the linguistically enhanced proposition that arises from it by requiring the first argument of identity to be cognitively identified via the term 'water' and the second to be identified via the term 'H_2O' is knowable only aposteriori. Those who believe this proposition, *understanding both* 'water' and 'H_2O', realize that assertive utterances of 'Water is H_2O' will typically assert that

the stuff, water, that comes in the form of a colorless, drinkable liquid that falls from the sky in rain is a chemical compound involving hydrogen and oxygen. Since speaker-hearers standardly presuppose that they understand the expressions, this informative proposition will normally be communicated and asserted, though, as we have seen, similarly rich descriptive enrichment will generally *not* occur under modal operators.

Here *understanding* the terms 'water' and 'H_2O' requires having different collateral information about what they stand for, despite the fact that their representational contents (which they contribute to the semantic contents of sentences containing them) are identical. A similar contrast can be drawn when understanding one term requires *recognitional ability* not required by understanding a second term with the same representational content. Let 'R' be a Millian kind-term designating the same surface spectral reflectance property that the color term 'red' does. Suppose that fully understanding 'red' (in the sense of knowing what is typically presupposed when it is used) requires being able to visually identify red things, whereas no such ability is required to understand 'R'. Then, the sentence, "the property *being red* is the property *being R*" can be used in a context in which it is presupposed that the term 'red' is fully understood to assert a proposition which—when combined with propositions represented by one's visual experience—allows one to draw informative conclusions that couldn't be drawn from the proposition semantically expressed by the sentence, which is the uninformative proposition that *the property being red is the property being red* (Soames 2015, 92).

Finally consider the English attitude ascription *Juan has just learned that water is H_2O* used to report a fact about a monolingual speaker of Spanish. Although the proposition semantically expressed by the ascription is, we many assume, false, the ascription can naturally be used to express a truth. This occurs when the semantic content of the ascription is enriched by requiring one who entertains the object of 'learn' to identify one argument of *identity* via the term 'water' *or some translation of it*, while identifying the other via 'H_2O', *or some translation of it*—where a term T_2 is a translation of T_1 only if conditions for *understanding* the two are roughly the same. Under these conditions, a use of the sentence *Juan has just learned that that water is H_2O* asserts that he has only recently come to believe a certain informative proposition that makes no claims about words

or translations. *When*, *as is pretty standard*, he is presupposed (i) to understand the relevant terms, and (ii) to take the descriptive information required by such understanding to be true of their referents, the assertive utterance will result in the assertion of a proposition that characterizes Juan as only recently coming to know that *that a certain stuff that comes in the form of a colorless*, *drinkable liquid and falls from the sky in rain is a chemical compound involving hydrogen and oxygen*.

This is just a sample of how getting the metaphysics of propositions right can help us make empirical progress in accounting for the semantics and pragmatics of hyperintensionality. Other examples include Kripke's Puzzle about Belief (Soames 2015, 81–84), puzzles about all manner of indexicals (Soames 2015, chapter 3, pp. 93–95, 112–16), and perceptual versions of Frege's puzzle (Soames 2015, 97–105).

References

Benacerraf, Paul. 1965. "What Numbers Could Not Be." *The Philosophical Review* 74 (1): 47–73. https://doi.org/10.2307/2183530

Boolos, George. 1984. "To Be Is to Be the Value of a Variable (of To Be Some Values of Some Variables)." *Journal of Philosophy* 81 (8): 430–49. https://doi.org/10.2307/2026308

Church, Alonzo. 1954. "Intensional Isomorphism and Identity of Belief." *Philosophical Studies* 5: 65–73. https://doi.org/10.1007/BF02221771. Reprinted in (Salmon and Soames 1988).

Church, Alonzo. 1988. "A Remark Concerning Quine's Paradox about Modality." In *Propositions and Attitudes*, edited by Nathan Salmon and Scott Soames. Oxford: Oxford University Press. Spanish version published in *Analysis Filosofico* 2, 1982.

Frege, Gottlob. [1884] 1950. *The Foundations of Arithmetic*. Trans. John L. Austin. Oxford: Blackwell.

Frege, Gottlob. [1893] 1952. "On Sense and Reference." In *Translations from the Philosophical Writings of Gottlob Frege*, 56–78. Trans. Peter Geach and Max Black. Oxford: Blackwell.

Gomez-Torrente, Mario. 2019. *Roads to Reference*: Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198846277.001.0001

Kripke, Saul. [1979] 1988. "A Puzzle About Belief." In *Meaning and Use*, edited by Avishai Margalit, 239–83. Dordrecht: D. Reidel. Reprinted in (Salmon and Soames 1988).

Maddy, Penelope. 1980. "Perception and Mathematical Intuition." *The Philosophical Review* 89 (2): 163–96. https://doi.org/10.2307/2184647

Mates, Benson. 1952. "Synonymity." In *Semantics and the Philosophy of Language*, edited by Leonard Linsky, 111–36. Champaign Ill: University of Illinois Press.

Perry, John. 1977. "Frege on Demonstratives." *The Philosophical Review* 86 (4): 474–97. https://doi.org/10.2307/2184564

Perry, John. 1979. "The Problem of the Essential Indexical." *Noûs* 13 (1): 3–21. https://doi.org/10.2307/2214792. Reprinted in (Salmon and Soames 1988).

Salmon, Nathan and Scott Soames. Eds. 1988. *Propositions and Attitudes*. Oxford: Oxford University Press.

Soames, Scott. 2010. *What is Meaning?* Princeton and Oxford: Princeton University Press.

Soames, Scott. 2015. *Rethinking Language, Mind, and Meaning*. Princeton and Oxford: Princeton University Press.

Soames, Scott. 2016. "Yes, Explanation is All We Have." *Philosophical Studies* 173 (9): 2565–73. https://doi.org/10.1007/s11098-016-0636-0

Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. Trans. G.E.M. Anscombe. 3rd edition. New York: Macmillan.

Organon F 27 (3) 2020: 302–324
https://doi.org/10.31577/orgf.2020.27302
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Feyerabend's Alternative Theories within Goodman's Worldmaking

Nicole Fišerová

Received: 25 January 2019 / Accepted: 26 November 2019

Abstract: The main purpose of this paper is to compare two pluralistic approaches to knowledge, Goodman's theory of worldmaking and Feyerabend's methodological anarchism. It therefore examines firstly, the concept of world-versions, which according to Goodman create our worlds and at the same time are crucial for achieving a better understanding of reality; and secondly, the concept of alternative theories which are built upon pluralism and, according to Feyerabend, secure knowledge and make scientific progress possible**.** Feyerabend's concept has been rejected by many, seemingly for its lack of limitations. In line with this argument, I propose that based on the comparison of these two pluralistic approaches, the alternative theories can be understood as a part of worldmaking, for Goodman's theory has wider applicability since it encompasses not only science but also art. Furthermore, I suggest adopting Goodman's principle of rightness, the criterion of functionality in his worldmaking, as a criterion within Feyerabend's methodological anarchism when establishing the prevailing theory. It is to be expected that such a juxtaposition will uncover inconsistencies, in particular regarding boundless relativism and the vague terminology in both conceptions.

* University of West Bohemia
Department of Philosophy, Faculty of Arts, University of West Bohemia, Sedláčkova 19, 306 14 Plzeň, Czech Republic
✉ niciryan@kfi.zcu.cz

Keywords: Alternative theories; methodological anarchism; Nelson Goodman; Paul Feyerabend; philosophy of science; theory of symbols; worldmaking; world-versions.

1. Introduction

Nelson Goodman introduced world-versions as part of his worldmaking (Goodman 1978) when trying to explain that we have not only one world, but many. There may be various versions, creating various worlds. If a version does not function well for us, causing us to misunderstand what is going on around us, we know it must be changed. Our worlds then change depending on the time frame, information, a given symbol system, etc.

Paul Feyerabend presents alternative theories as a part of his methodological anarchism (Feyerabend 1975).[1] He advocates the idea that a plurality of theories guarantees the development of science, where one theory generates its alternatives immediately upon its accession. An alternative theory replaces the old one when it is no longer sufficient to explain phenomena.

Both conceptions show similar pluralistic features and further claim that there is no symbolic system or methodology to be preferred while describing the world. However, I believe that a preferred symbol system and set of rules must be established, particularly within scientific discourse, if science wants to retain its prominence. Feyerabend's alternative theories can therefore be interpreted not only within Goodman's worldmaking, but the criterion of rightness, which was originally designed by Goodman, could be used for Feyerabend's methodological anarchism when describing the process of theory selection.

According to both Goodman and Feyerabend, pluralism is a good starting point for any discourse. However, only pluralism itself can neither create versions with no criteria nor ensure scientific progress. It works well for Goodman because he realized the danger of an uncontrollable pluralism and therefore set other criteria to avoid absolute relativism. On the other hand,

[1] Methodological anarchism was first presented in *Against Method,* published in 1975; for the purpose of this study, the 1993 edition is used.

Feyerabend built his concept upon pluralism, for he tried not to contradict his own anarchism: yet to leave pluralism "taking care of the job" seems implausible if the concept of alternative theories is to be functional. Moreover, if one examines his concept one is left with the need for some form of criterion. Therefore, I suggest using Goodman's very own criterion of rightness, for it is flexible in definition (defined in relation to the specific subject or discipline) and leaves pluralism somehow unharmed.

2. Goodman's worldmaking

Goodman presents his theory of worldmaking, a conception of knowledge of the world expressed in various symbolisms, in *Ways of Worldmaking.* It embodies Goodman's claim that in the course of modern philosophy, the structure of the world has changed. From the initial seeking of such a structure in the reality surrounding us, we turned our research to the structure of mind, then to the structure of language, then finally to the structure of symbols. In other words, a fixed world which was supposed to be found was exchanged for the diversity of the several symbol systems of the sciences, the arts, philosophy, everyday discourse, and perception. He claims that worlds cannot exist without symbol systems; they are dependent upon them. Symbol systems are created by humans and they help us retain some kind of structure and order in the arts and sciences (Goodman 1978, x).

Goodman claims that there is no such thing as the real world—a ready-made, unique, independent, absolute reality, for "there are many worlds, if any." In his conception, the one world may be taken as many or the many worlds taken as one; it only depends on how one takes it (Goodman 1978, 2).

From Goodman's perspective, pluralism about literature appears more plausible than pluralism about reality. The proposition declaring that different interpretations define different worlds looks much more believable and clearly less dubious than the proposition that different versions define different worlds,[2] although Goodman and Elgin (1986, 567) claim that there

[2] From whence follows Goodman's conviction that a single text underlies different interpretations. That may seemingly, by analogy, support the conclusion that a single world underlies different versions.

are important differences between the two cases worthy of further discussion.

These worlds should not be conflated with possible worlds; they are all actual (Goodman 1983, 271). They are made of the so-called "world-versions,"[3] of which some can be irreconcilable and in conflict with others (Goodman 1978, 3). Versions are perceived under one or more frames of reference; we cannot say anything about the world in itself apart from the all frames.[4] Versions are made of various symbols, and they may encompass descriptions, depictions, pictures, world perceptions, etc. Even a point of view can be considered to be a world-version (Goodman 1978, 5).

Goodman finds it easier to talk about versions rather than worlds themselves, possibly because it has never been clearly stated how many versions one world can have. However, if one asks about the content of the worlds, he explains it but, by doing so, denies any solid foundation. There are many stuffs—matter, energy, waves, phenomena—that arise along with the worlds. He goes even further. He clarifies that worlds are never made from scratch, for worldmaking starts from the worlds already known. Therefore, making is remaking; it is a process of building a world from others. There is always some old version or world[5] that we have at hand; we are stuck with it until we have the determination and skill to remake it into a new one. In other words, worldmaking begins with one version and ends with another (Goodman 1978, 6–7, 97).

Worldmaking does not stop there, though. After having accepted the proposition that a world is made by worlds that are but versions, with substance dissolved into a function, one must face the questions of how worlds are made and tested. Goodman suggests the following ways[6] of making a world: a) composition and decomposition, b) weighting, c) ordering,

[3] Sometimes plainly referred to as "versions."

[4] Goodman (1983, 270) expresses the same thought metaphorically: "The innocent eye is a myth long dead."

[5] Goodman (1978, 97–101) admits that philosophers from antiquity, such as Thales, Anaximander and Democritus, had their world shaped by religion, superstition, suspicion, hope and experience.

[6] The definite article is here omitted intentionally. Goodman only suggests possible ways. The classification should not be taken as mandatory or clearcut.

d) deletion and supplementation, and e) deformation. These processes often occur in combination. As for testing, it will be discussed in the following section (Goodman 1978, 7–17).

3. Goodman's world-versions

Goodman takes relativist thinking into account but radical relativism has no place in his philosophy. There are many versions, but it does not mean that all of them are right; many can even contradict each other. However, everything we can learn about the world is contained in its right versions. For better understanding, we may want to define the relation among them and sort them into clusters, each cluster constituting a world. For many purposes, though, we can simply use the term 'versions' for the ways-the-world-is (Goodman 1978, 4).

Versions, as mentioned above, are made by any kind of symbols. There can be such containing words, numerals, pictures, sounds, or other symbols of any kind in any medium. However, regardless of their similarities or differences, Goodman considers the comparative study of the versions and visions and of their making to be a critique of worldmaking, for he claims that versions as such are incommensurable. What is meant is that in some cases, we cannot claim a verbal version to be better than a visual one if both bring the same piece of valuable information. The wording here is crucial. We can say that one of them is preferred or more suitable for a particular circumstance, but not that *this* is right and *that* is wrong (Goodman 1978, 94).

Therefore, the versions we create can be further segmented into literal, non-literal, metaphorical, verbal, non-verbal, etc.[7] There is no preferred version, therefore even no preferred symbolic system (language) which we use for describing the world. Whence it follows that there is no reason to prioritize scientific descriptions on the grounds that science is believed to provide us with neutral facts,[8] in contrast with the vague and metaphorical arts

[7] Goodman never gave a full list of possible versions. The above were collected from (Goodman 1978) and (Goodman and Elgin 1988).

[8] The notion of science as a collection of neutral facts.

(non-science). Science and the arts have the same main objective, which is knowledge (Goodman 1978, 1–6).

According to Goodman, the reason why we keep on preferring some versions, particularly the scientific, lies in convention. Some theories, hypotheses and predicates are prioritized while some others are left behind. It is, however, only a matter of habit and experience that we believe that all emeralds are green rather than *grue* (Goodman 1955). Even if we consider the entrenchment[9] of a predicate, Goodman's very own attempt to solve his New Riddle of Induction, we find ourselves stuck in a circle. However, entrenchment plays a great part; it derives from the use of language and results from the actual projections conducted in the past. Therefore, it leaves us dependent on convention, past experience and congruence with practice and the actual use. As a matter of fact, Goodman himself does not consider his the only possible solution and later realizes the importance of experience and habitual action, which can be understood as the practice of language users. Putnam even states that Goodman finds entrenchment not to be innate but resulting from philosophical reflections regarding the practice of a language community (Goodman 1955, viii). In *Ways of Worldmaking*, he speaks of rightness of categorization instead, and admits that it is a matter of fit with practice (Goodman 1978, 138).

As a result, he claims that we simply prefer to apply to a sphere of objects a specific scheme which we find comprehensible. The issue is that science gives us facts, and facts are problematic because each version states its own facts (Goodman and Elgin 1988, 125, 183–84).

The vast variety of versions is striking not only in the sciences, the works of various painters and writers, but also in our perceptions as influenced by circumstances, by our own insights and past experiences. We can have contradictory versions which may be right in different systems, but that does not mean that all versions are right; it is important to distinguish between

[9] Generally speaking, if a term or a predicate is entrenched it has an established position in our language practice. Goodman's theory of entrenchment originally appeared as an attempt to solve his own *Grue* paradox. Its principle is based on ordering hypotheses in the light of past inductive practice. It depends upon the record of past actual projections and the frequency with which the predicates were actually inductively projected in the past (Goodman, 1955, 84, 94–95).

those that are right and those that are wrong. We cannot just test a version by comparing it with a world undescribed, undepicted, unperceived (Goodman 1978, 3–5). The process of telling the wrong version from the right one is more complicated than that; it is rather a matter of interaction between symbol users and the assumed world. Every version is further tested and confirmed by rightness (Goodman 1978, 109–38).

4. Feyerabend's epistemological anarchism

A similar pluralistic approach to knowledge is offered by Feyerabend when he comes with his epistemological anarchism or also methodological anarchism. It serves as a critique of methodological monism which, he claims, does not lead to any progress in science or its development. This ironic stance was described in *Against Method*. Drawing upon various examples from the discourse of science, he shows how irrational it would be to claim that only one correct scientific method could grant us progress (Feyerabend 1993, xiii–xiv).

Feyerabend's conception refuses to ascribe the priority to scientific descriptions just on the grounds that science is believed to provide us with neutral facts. Science is, in its very essence, an anarchic enterprise. Theoretical anarchism is, thus, more humanitarian and more likely to encourage progress (Feyerabend 1993, 39). Needless to say, Feyerabend's stance is not aimed against science so understood[10] but against ideologies that use the name of science for "cultural murder." Cultural murder is committed when the progress of knowledge involves killing minds that is, according to Feyerabend, connected to the process of pushing Western ways and values into all four corners of the globe. Generally speaking, the "killing of minds" can be understood as the rejection of non-scientific procedures, as it is

[10] "I am not against science. I praise its foremost practitioners and (in the next section) suggest that their procedures be adopted by philosophers. What I object to is narrow-minded philosophical interference and a narrow-minded extension of the latest scientific fashions to all areas of human endeavor—in short what I object to is a rationalistic interpretation and defence of science." (Feyerabend 1993, 122)

believed that results gained by means of non-scientific methods are not to be taken seriously (Feyerabend 1996, 3–4, 14).

For we have no unified scientific method that contains unchanging and absolutely binding principles for conducting science; the belief that science is the best way of gaining knowledge proves to be unjustifiable, and, moreover, the procedures and results that constitute the sciences have no common structure (Feyerabend 1993, 1). The results presented by science do not alone prove its excellence since they often depend on the presence of non-scientific elements, and such elements, points of views or methods are both necessary and beneficial to science. Therefore, one must accept that science contains not only one but many approaches to research (Feyerabend 1996, 26). Today science prevails, according to Feyerabend, not because of its comparative merits, but "because the show has been rigged in its favor" (Feyerabend 1978, 102). Thus the biggest issue is that science is supposed to be about something while creativity need not be (Feyerabend 1996, 24).

It has been clearly stated that the whole idea of a fixed, unified method or even a fixed theory is naïve and maybe a little preposterous. If we want to keep any objectivity, precision or truth, there is only one principle which can be defended under all circumstances and does not inhibit any scientific progress: *anything goes* (Feyerabend 1993, 18–19).

From this perspective, pluralism, and therefore a pluralism of methods, affords us the best chance of securing knowledge. We shall ask for the freedom of science and its free revolutions because only a plurality of theories ensures scientific progress and only by means of such progress do we gain knowledge (Feyerabend 1999, 4–5).

5. Feyerabend's alternative theories

Feyerabend aims his research primarily at scientific methods and their descriptive apparatus. However, he speaks mainly of science not because he finds other aspects of life to be less important but because he realizes that people tend to set borders between science and non-science and attach more importance to science due to its sometimes illusory credibility. He does not think it is special. We have no strict sets of rules in life; we may have laws and moral codes but no one can really tell us how to live. However, in

science, thinkers and philosophers have been trying to set such rules and tell scientists how to do their job. Pluralism is taken as part of everyday life, disguised as freedom or tolerance; it is much more complicated to accept pluralistic views in science; hence Feyerabend's insistence.

Moreover, it seems that in an ideal case we would have similar criteria in both life and science, i.e. freedom, passion and pluralism leading us to a better world. Feyerabend's idea of science that regulates itself is tempting; however, life does not need to be regulated for it is not perceived as a hard science.

Feyerabend criticizes the concept of science as a symbolic system that describes neutral truths and independent facts. Inevitably, with the critique of one unified method for science comes the critique of one prevailing theory that can embrace and interpret all facts. He believes that no theory ever agrees with all the facts in its domain, but yet not always it is the theory that is to be blamed, for facts are constituted by older ideologies, and there is a reason to suppose that a clash between facts and their theory may be proof of progress. Besides, if a theory clashes with evidence, the reason may be that the evidence has been contaminated by wrong samples or wrong measurements (Feyerabend 1993, 39).

As mentioned above, Feyerabend believes that a plurality of methods grants us progress in science and, therefore, gives us knowledge. In the light of this thought, he further asks for the freedom of science and its free revolutions, since a plurality of theories should guarantee free scientific progress (Feyerabend 1999, 4–5). In the history of science, people have always been pursuing unity,[11] whether it was unity in methodology or a unified theory rich enough to produce all the accepted facts and laws (Feyerabend 1993, 43).

His thesis about the influence of theories on our observations criticized the legitimacy[12] of observation statements and was supported by the claim

[11] According to Feyerabend, this desire for unity that underlies the many events surrounding us comes from the Western sciences (Feyerabend 1993, 43).

[12] As a thinker inclining to constructivism and taking Goodman into consideration, I would even speak of the *validity of observational statements,* for I believe that the validity of statements can be determined by referring to particular observations. For a non-constructivist, however, that may seem ineligible or even categorically wrong.

that observations (observation terms) are not only theory-laden but in fact fully theoretical (Feyerabend 1981, x–xi). Therefore, one must consider that observational reports or 'factual' statements either contain theoretical assumptions or assert them indirectly by the manner in which they are used, from which it follows that all facts are theoretical. Consequently, facts are both influenced and constructed by the prevailing theory. Furthermore, neither the rules, nor the principles, nor even the facts are sacred—we may, therefore, change them or even create new facts and new grammatical rules, and see what happens once these rules are available, applicable and have become familiar. One must note, however, that such an attempt may take considerable time (Feyerabend 1993, 22, 123).

The whole conception of alternative theories, where the prevailing theory generates its alternatives that sooner or later take over, has evolved from the thought mentioned above: *pluralism ensures scientific progress*. The progress is, however, kept alive thanks to the so-called scientific cycle, which is described in the following manner: scientific revolutions ensure a new theory and the new theory generates its alternatives immediately upon its accession. Therefore, such a plurality of theories that are both in conflict with each other and incommensurable[13] should then present crucial elements for maintaining the advance of science (Feyerabend 1993, 152–55).

Feyerabend gives a further exemplary explanation. He claims that research always starts with a problem, which results from a conflict between expectation and observation. However, an observation is constituted by the expectation. After having formulated the problem, one can start solving it—finding a theory that is feasible, relevant and falsifiable, but not yet falsified. In the next step, one has to criticize the theory that has been put forth when attempting to solve the problem. If successful, the criticism will eliminate the theory for good and simultaneously will create a new problem. If one wants to solve the newly arisen problem, one needs a new theory that reproduces the successful consequences of the older theory, denies its mistakes and makes additional predictions not made before.

It follows that there are two theories which overlap, the first being the old theory, the second one the new theory and the intersect represents the

[13] Feyerabend borrows the term "incommensurability" from (Kuhn 1962), but uses it in a distinctive way.

problems and facts of the old theory that are still remembered and have been distorted so as to fit into the new framework. This cycle keeps repeating itself. A theory generates its alternative immediately upon its accession and prevails until it is replaced by its alternative. Such a proliferation of incommensurable theories and their conflicts should keep science advancing, thus bringing us knowledge. In the light of this, it can be claimed that knowledge is not a series of self-consistent theories. It is not a gradual approach to the truth either; rather, it is an ever-increasing ocean of mutually incompatible alternatives (Feyerabend 1993, 152–57, 22).

6. Similar features in Goodman's and Feyerabend's approaches

There never was a discussion between Goodman and Feyerabend concerning multiple actual worlds or alternative theories. The most famous interaction was Feyerabend's reflection of Goodman's "new riddle of induction." Feyerabend's interest in Goodman's paradox was in accordance with Feyerabend's critical and deprecatory approach to rule-following in science; it shows how substantially Feyerabend was influenced by the later Wittgenstein.[14]

However, I argue that there are many similarities between the concepts that are worthy of further investigation. The most significant, which we shall examine individually, are that both theories 1) have a pluralistic background, 2) refuse to take truth as the main criterion for testing, 3) have a similar language-shaping interpretation of reality, 4) claim that sciences influence or even fabricate facts and, therefore, question the superior position of science when gaining knowledge, and last but not least, 5) consider our knowledge/understanding of the world to be always more or less partial.

Both Goodman and Feyerabend advocate pluralism over monism, but needless to say from different standpoints. Goodman goes beyond scientific discourse; he requires pluralism in all possible domains. In his conception,

[14] For more detailed analysis regarding the analogy between Goodman's and the Wittgensteinian criticism of inductive reasoning see (Schuster 2018); concerning Wittgenstein's "profound" influence on Feyerabend see (Feyerabend 1978, 114).

there can be a version which is scientific or even a world made of various theories/versions; however, as we know, not all the versions are scientific, there are many of them and they are made of many kinds of symbols.[15] Therefore, he not only asks for the plurality of versions but he also claims that there are many worlds. There is no world "w," no absolute reality waiting to be revealed by science. The whole idea that there is knowledge completely independent of the observer and neutral in every way is direly suspicious (Goodman 1978, 131–32).

Feyerabend is against monism in science and demands a plurality of methods within scientific frames, where any method can be right under specific circumstances. He believes that if a scientist wishes to understand the empirical content of our views as good as he can, he must introduce and use other views and adopt a pluralistic methodology (Feyerabend 1993, 14, 21). The pluralism which Feyerabend favors involves taking a wide variety of different methods and accounts of the world into consideration (Lloyd 1996, 252). Based on that he asks for a plurality of theories, which should guarantee free scientific progress, for he believes that actual science is much closer to pluralism than the defenders of monism would like to admit (Feyerabend 1981, 111).

His claim is further supported by the explanation that knowledge needs a plurality of ideas, even non-scientific ones, and by that it can be proven that well-established theories are never strong enough to eliminate alternative approaches. Such various approaches create various theories, thus pluralism is practically inevitable (Feyerabend 1993, 131). However, more than that, it is a part of our every-day life because "there are many ways of thinking and living" (Feyerabend 1995, 164).

Along with adopting pluralism as an important part of their conceptions, another issue comes into question. So much has been said about various versions and visions, about theories and their alternatives, but thus far no criterion has been set in order to distinguish those that are right from those that are wrong. How can we test or prove any to be wrong if for both Goodman and Feyerabend it is impossible to prove a version/theory to be wrong by comparing it with the world accessible to us?

[15] The general theory of symbol is presented in (Goodman 1968).

When searching for the objectives and the constraints of worldmaking, Goodman opens this topic and tries to set the criteria for success in making a world. It is clear that different standards are needed for understanding the vast variety of versions. For example, the distinction between true and false works well for us if applied to descriptive statements or scientific versions, but falls short if applied to metaphorical statements or paintings. Worlds, however, can be presented and made in many ways—in scientific theories, works of art, and versions of many kinds. Worldmaking goes far beyond theories, descriptions or statements, even language; it involves all kinds of versions. Goodman's relativism allows a greater number of right versions, but that does not mean that we make a world by putting symbols together at random. We must tell a right version from a wrong one by means of a criterion which, for worldmaking, is rightness. A version is not so much made right by a world as a world is made by a right version (Goodman 1978, 109, 94).

A consideration of standards other than truth is necessary, considering that we encounter not only literal statements; versions that make no statements have to be included as well. This leads us to the conclusion that truth is often inapplicable, hardly sufficient, and must sometimes give way to competing criteria.[16] Therefore, the true/false criterion seems insufficient for considering versions in general. Goodman offers rightness to be more adequate; it has wider application, considers temporality and is shaped and formed by circumstances (Goodman 1978, 107). Truth can be, of course, an occasional component of rightness (Goodman and Elgin 1988, 181).

Goodman admits that the conception of truth appears to be adequate for science but beyond science we do not always seek truth. Pictures or melodies are not considered to be true or false. The conception of rightness is proposed as a criterion in non-sciences, claiming that rightness of design differs from rightness of representation or description not so much in nature

[16] Goodman explicitly states: "Some truths are trivial, irrelevant, unintelligible, or redundant; too broad, too narrow, too boring, too bizarre, too complicated; or taken from some other version than the one in question, as when a guard, ordered to shoot any of his captives who moved, immediately shot them all and explained that they were moving rapidly around the earth's axis and around the sun" (Goodman 1978, 120–21).

or standards as in the type of symbolization and the mode of reference involved. Hence it follows that the truth of statements and the rightness of descriptions, representations and exemplifications is primarily a matter of fit, or fit with practice (Goodman and Elgin 1988, 136–39).

In sum, truth can be easily reformulated as the rightness of a true or false statement. Therefore, it seems intelligible to claim that rightness has a wider use—not only in the arts and non-sciences but rightly formulated, even in science.

Although it may seem complicated to abandon truth as a criterion according to which we test our knowledge, and adopt the conception of rightness both in science and the non-sciences, it may be even harder to live with no criterion whatsoever. If we have a closer look at Feyerabend, the conception of alternative theories seems rather difficult to adopt if we lack a standard according to which we decide which theory prevails. He claims that no theory can be refuted by means of confrontation with empirical facts. Yet for him, truth equals fact; these two terms can be interchangeable. After having disposed of truth, we may logically turn to consensus. However, according to Feyerabend, consensus[17] is deadly for the development of knowledge because it retards scientific progress (Lloyd 1996, 257). If we cannot take truth for a criterion and we do not have any other, and if we cannot rely on consensus, it seems impossible for us to decide which theory wins and why.

Although Goodman finds truth to be insufficient, he still somehow keeps it in business where science is considered or when speaking of literal statements; although masked by rightness, we can still find truth to an extent important in this field. Contrasting with Goodman, Feyerabend's field of study was restricted to science and yet he finds truth unimportant when refuting a theory because each prevailing theory states its own truth, or if we may so call it, *truthness*. What seems important to both Feyerabend and Goodman are functionality and consensus, although the latter is understood differently by them.

Goodman understands rightness as "standards of acceptability that sometimes supplement or even compete with truth where it applies, or even

[17] Nevertheless, Feyerabend realizes the power of consensus, however deadly he himself may consider it.

replace it for non-declarative renderings" (Goodman 1978, 110). Functionality and convention then create rightness. It needs to be pointed out, however, that it refers to the functionality of a system as a whole, after one adds a piece or pieces of information, not to a piece of information functioning by itself. Rightness, according to which some statements or depictions are proclaimed to be valid, can then be understood as a matter of the habitual action and practice of the symbol users (Goodman and Elgin 1988, 183).

Feyerabend, on the other hand, described the whole process of an alternative theory being somehow "born" from the old one, but the scientific cycle where one theory beats the other seems to be more important to him than finding a key, a criterion by which one theory ceases to be sufficient for scientists and knowledge seekers. Therefore, I argue that Goodman's conception of rightness could serve as a criterion for how each theory is selected to become the prevailing one. I further suggest that using such a criterion will show that Feyerabend's alternative theories may be partly feasible, but have a lot of inconsistencies. Moreover, it is necessary to highlight that the complete rejection of truth is not an option either in science or society because it is supported by our habits and the cultural environment in which we live.

Feyerabend insists that his main purpose is neither to substitute one set of rules with other sets of rules nor to offer some new standards, a new methodology which need be followed in science; his intention is to show that all methodologies have their limits (Feyerabend 1993, 32). However, his very own philosophy of science shows that no matter how pluralistic we may be and how many rules his philosophy allows us to break, not really everything goes. It exemplifies the truth that not only the dogmatic but also the anarchist methods have limitations.

In the light of this, it is also necessary to show Goodman's and Feyerabend's similar views on language. Goodman, with his constructivist attitude,[18] claims that there can be a language without worlds but no world

[18] Goodman avoids labelling and, usually, any kind of generalization; however, in *Reconceptions in Philosophy and Other Arts and Sciences* he clearly inclines towards constructivism, although he admits that it still needs a lot of work (Goodman and Elgin 1988, 189).

without words or other symbols. We then interpret what we perceive and shape our worlds accordingly. There is no world by itself, independently of language or symbol system. Such systems are created by people and they help us give structure and order to art and all the sciences (Goodman 1978, 6).

Feyerabend thinks similarly. He refers to Whorf, who had earlier formulated the theory that languages and the reaction patterns they involve are not just instruments for describing events, facts or states of affairs, but also shape them. The observation language, which we adopt with an alternative theory, then logically influences its facts (Feyerabend 1993, 164).

Both authors have, therefore, no difficulty in admitting that humans more or less co-create our reality; they are part of this world-construction. However, possibly for this reason, they also refuse to take the priority of scientific facts or the prior position of science in general.

Goodman points out the misleading power of perception and argues that not only science fabricates facts, but even perception makes its own. He finds it pointless to believe that facts are found, not made, and that they constitute the one world that is to be revealed by scientists (Goodman 1978, 89). On that basis, he further refuses to take the prior position of science and the "neutral" facts that are presented to us by scientists. He states that science denies its data and picks the right samples, which are then presented to us. Each theory has to be adjusted to fit the facts as much as facts have to be adjusted to fit a theory (Goodman 1968, 263). The dominant position of science against art is therefore unjust because facts are made as well as our worlds are. For Goodman, it is all about making and remaking; art and science have the same goal, which is knowledge. More precisely, they grant us better understanding.

Feyerabend's position is very clear on this matter, for as we have no unified method we cannot justify the preeminent position of science in our society. Science is just one way of gaining knowledge or information and not necessarily the best one. He adds that neither science nor rationality are universal measures of excellence; they are more likely particular traditions (Feyerabend 1993, 214).

He finds it problematic that many "educated citizens" take it for granted that reality is only what science or scientists say it is and that beyond that,

other opinions may be recorded, but need not be taken seriously. This picture of science is, however, wrong because he believes that science offers not one story, but many (Feyerabend 2001, 27).

Not only science offers many stories; for Goodman, many stories are presented by all versions, which include not only science but even non-science. With each right version being part of our world, we come to a better understanding of it.[19] Needless to say, our understanding is always partial. It comes with the seemingly endless cycle of making and re-making the world or worlds (Goodman and Elgin 1988, 161–62). Here we can see an obvious similarity between Goodman and Feyerabend. Successful worldmaking, which is to be achieved by getting to know the right versions and which is therefore always more or less incomplete, matches Feyerabend's alternative theories conception, for each theory brings new information and new views, leading us closer to knowledge. Both agree that our knowledge or understanding can theoretically be complete but that in practice we neither have the time nor the capacity to make it so. In summary, we eventually wind up with partial knowledge and partial understanding. Although it may seem reasonable to claim that the former and the latter can be interchangeable, the next section is intended to show that there is a reason for keeping them apart.

7. Different features of Goodman's and Feyerabend's approaches

Aside from all the similarities between both authors' theories, there are some asymmetries to be examined. In both theories, there are 1) different discourses of examination and different principles used, as noted above, and 2) several difficulties with distinguishing between incommensurability and comparison. Although these differences are few, they may prove crucial. Different discourses and the lack of a criterion are the very reasons for

[19] For Goodman and Elgin, understanding has a wider range of use than knowledge, although knowledge can be a part of understanding. There we can see a similarity with the conception of truth and rightness.

interpreting Feyerabend's methodology as a functional part of Goodman's theory of worldmaking.

As already mentioned, according to Goodman, science is only one of many accessible symbolic systems by means of which we create so-called versions. However, our world consists not only of verbal and literal versions, but of various kinds. In order to gain a greater understanding of things around us, ideally we need to get to know all the right versions and make sense of them.[20] However, Goodman admits that although we should try to make sense of them, it seems improbable that a human being should reach such a state of complete understanding/knowledge. If we add a piece of information to "our" version, it will never make sense by itself; a version works like a whole system. If the whole system works and if it further corresponds with our other beliefs, points of view and so on, we have made the version better through addition. Such a piece of information can be basically anything: a new element, a particle, a belief, even a book, but most importantly for scientific discourse, a theory.[21] Goodman's conception goes far beyond the discourse of science. As he points out, his primal concerns in worldmaking are metaphorical versions—worlds of fiction, poetry, painting, music, dance, etc., for he finds non-scientific discourse and its versions to have been rather neglected. However, he is willing to take "the real world to be that of some one of the alternative right versions (or groups of them bound together by some principle of reducibility or translatability) and regard all others as versions of that same world differing from the standard version in accountable ways" (Goodman 1978, 20).

In contrast, Feyerabend stays within the scientific discourse. When talking about his alternative theories, he applies his conception only in the realm of science and describes the ideal progressive theory cycle. However, he does claim that non-scientists are indispensable for scientists, for the non-scientific element is crucial for further development in science; yet this area lingers logically unexplored within his research. Goodman expanded his theory thanks to his general theory of symbols; Feyerabend did not have to cross this border when dedicating his research to the methodology of science, for he only used one symbolic system (language of science). It is

[20] The world is what all the right versions are.

[21] A theory can be either a version by itself or just part of a cluster of versions.

therefore understandable that Goodman needed a criterion with wider application of which the occasional part may be truth (for science), but Feyerabend dispenses with truth because science fabricates facts. These are further given to us when we accept a theory; in other words, facts are created by older ideologies and a specific theory itself. Pushing this line of considerations into extremes, it would mean that a scientist is able to predict all the facts which the prevailing theory is able to comprehend in a given domain.

When the theory proves to be insufficient, which means that it is no longer able to explain all physical events, we accept its alternative. After some time, once we are no longer able to predict more facts, an alternative takes over and this cycle keeps on repeating. The issue with such cyclic progress is clear: who is then to state according to which criterion we accept facts that are presupposed and predicted by a theory? It is certainly neither the state nor the church, for science should be independent of both of them (Feyerabend 1993, 39).

In such a cycle, Feyerabend proposes two main principles which should support scientific progress and lead to it: counterinduction and proliferation. Counterinduction is then a legitimate and reasonable move in science. Perhaps the best example of the principle of counterinduction would be Feyerabend's metaphorical comparison in which he claims that "we need a dreamworld in order to discover the features of the real world we think we inhabit," since the world cannot be explained from the inside by the principle of induction but by means of an external standard of criticism (Feyerabend 1993, 53, 22). If Goodman's concept embraces all versions, it should be possible to explain Feyerabend's stance within the frames of worldmaking. An alternative theory of a prevailing theory could be explained as a remaking of our old version, for we never start a version from scratch. However, a dreamworld Feyerabend describes could then be considered as a counterfactual version. It would not meet the required condition of being functional, but by showing a non-functional version we can test which ones work well for us.

In the light of his previous thoughts about facts being made, Feyerabend takes a stand that no theory can be refuted by means of confrontation with empirical facts. In other words, facts are created by older ideologies.

Goodman wrote a whole chapter about facts being fabricated and how the right exemplification can shape the sphere of objects within a theory. In other words, no version can be proven wrong by comparing it to the world accessible to us. Accordingly, no version can be refuted by means of confrontation with empirical facts (Goodman 1978, 91–92). However, the issue seems to be that Feyerabend postulates only one world described and revealed by means of many ungoverned methods and theories, whereas Goodman supposes many ways of worldmaking for many actual worlds.

Regarding this topic, Feyerabend further believes that a scientist must compare ideas with ideas rather than with "experience." Theories can be compared but the condition of the incommensurability of theories has to be taken into account. Each theory uses different observation languages and even if an identical term appears in two of them, they may differ semantically, so theories are incommensurable (Feyerabend 1993, 21, 51), whereas Goodman claims that the comparative study of versions and visions and of their making is a critique of worldmaking. It then follows that one should not compare versions with one another, because should they bring us the very same understanding/knowledge by different means or using different methods we would not be in a position to decide which way was the more eligible. This especially regards phenomena presented by science—a way which is, according to Goodman, unjustly prioritized—and phenomena depicted or performed by art (Goodman 1978, 94).

8. Conclusion

Considering what has been written about both theories, it seems peculiar but possible to consider Goodman's worldmaking to be an open, wider theory with no limitations in applicability and application; and Feyerabend's alternative theories within his methodological anarchism to be a closed one.

Goodman assumes that we co-create our world; we make versions. His goal is not only knowledge, which is sought by science, but understanding, which is typically favored by non-sciences. For better illustration, it could be described as a domain of understanding which encompasses the domain of knowledge. It becomes even more complicated if we imagine that we may

have uncountable amounts of domains of understanding containing smaller domains of knowledge, for we have many actual worlds.

Feyerabend, on the other hand, describes the domain of knowledge, for he particularly focuses his research on science and its methodology. However, the question remains: how is it possible for both concepts to have so many similarities and critical features when the field of application (one world in Feyerabend's, many worlds in Goodman's) is so different?

Let us start with an exemplification of Goodman's rightness working within Feyerabend's alternative theories as suggested at the beginning of this study. If we try to apply rightness to Feyerabend's concept we encounter two issues with which we need to deal. Firstly, we may inductively use the criterion of rightness in individual cases and secondly, we should come to realize that the criterion can be used even on a larger scale when setting a prevailing theory. In other words, it is necessary to narrow down the field of examination and the extension of rightness itself.

Feyerabend's famous example of Galileo, who built his theory upon an *ad hoc* hypothesis, was supposed to show not only that we do not need a unified method but also that having it would prevent progress in science. Instead, Galileo identifies the natural interpretations that contain an idea of the relativity of all motion and the law of circular inertia, and creates a new observation language (Feyerabend 1993, 54–55). If we apply the criterion of rightness defined as the functionality of a system as a whole in a conventional system of symbols[22] we discover that Galileo's idea works well for us when acting on our reality and coping with it.[23] Had he followed the notion of reality required by the church, society, or even scientists obeying given rules and methods, never he would have made such an important discovery.

As *ad hoc* as his hypothesis might have been, its functionality justifies the means by which it was acquired. However, that is why I argue that unknowingly—for the term was not coined back then—Galileo acted in accordance with the criterion of rightness all along. Despite his hypothesis having been built on a spontaneous basis, he did follow the rules defined by

[22] Yet we must not think of rightness as a convention or habit in the literal sense.

[23] The terms "acting on the world" and "coping with the world" are borrowed from (Dreyfus and Taylor 2015).

rightness. He discovered that his hypothesis worked well for us; if it had not, then the whole idea would have been lost and thus ultimately not taken seriously in the scientific discourse. On the same principle we can decide which theory is prevailing and which is already outdated.

Having shown the practical use of rightness, and thus demonstrated its use in scientific discourse, it appears to give rise to another question: who will define the criterion of rightness? I would argue that the simplest answer would be: scientists. They should be responsible for the definitions of individual cases, for they should be sufficiently qualified and able to defend their own methods and uses. A more complex answer would be: scientists under the supervision of philosophers.

We arrive at an interesting, unexpected conclusion: Feyerabend strictly insists on not following any dogmatic set of rules or any methodology, but nevertheless advises replacing induction with counterinduction, which contradicts his initial intention of not offering any new set of rules. Furthermore, it seems that in his pursuit of anarchy he unavoidably set some rules, for preaching "anything goes" sets a limitation if it excludes the possibility of having only one right method. This entails that science must have some rules, some regulated methodologies; however, what it does not show is who gets to do the choice.

It has been supposed that the concept of rightness could serve as the criterion for selecting alternative theories. Thus I argue that it would be possible to interpret Feyerabend in the context of Goodman's pluralism. The concept of alternative theories can be considered as part of Goodman's worldmaking, representing only the scientific domain, which is governed by knowledge. Therefore, there will be many alternative and incommensurable theories competing with each other and by such means, science will advance and our knowledge of the world will deepen. Outside the sphere of science, there will be the domain of understanding, where the main purpose will be to determine, by means of rightness, the functionality of various versions.

Funding

This paper has been supported by the project ZČU SGS-2017-031.

References

Dreyfus, Hubert, and Taylor, Charles. 2015. *Retrieving Realism*. Cambridge, MA: Harvard University Press.

Feyerabend, Paul. 1978. *Science in a Free Society*. London: NLB.

Feyerabend, Paul. 1981. *Realism, Rationalism and Scientific Method: Philosophical Papers Volume 1*. New York: Cambridge University Press.

Feyerabend, Paul. 1993. *Against Method*. 3. ed. New York: Verso.

Feyerabend, Paul. 1995. *Killing Time: The Autobiography of Paul Feyerabend*. Chicago: University of Chicago Press.

Feyerabend, Paul. 1996. "Theoreticians, Artists and Artisans." *Leonardo* 29 (1): 23–28. https://doi.org/10.2307/1576272

Feyerabend, Paul. 1999. *Knowledge, Science, and Relativism: 1960–1980*. Edited by John Preston. New York: Cambridge University Press.

Feyerabend, Paul. 2001. *Conquest of Abundance*. Chicago: University of Chicago Press.

Goodman, Nelson. 1955. *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Goodman, Nelson. 1968. *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis: The Bobbs-Merrill Co.

Goodman, Nelson. 1978. *Ways of Worldmaking*. Indianapolis: Hackett Pub. Co.

Goodman, Nelson. 1983. "Realism, Relativism, and Reality." *New Literary History* 14 (2): 269–72. https://doi.org/10.2307/468685

Goodman, Nelson, and Elgin, Catherine Z. 1986. "Interpretation and Identity: Can the Work Survive the World?" *Critical Inquiry* 12 (3): 564–75.

Goodman, Nelson, and Elgin, Catherine Z. 1988. *Reconceptions in Philosophy and Other Arts and Sciences*. Indianapolis: Hackett Pub. Co.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lloyd, Elisabeth A. 1996. "The Anachronistic Anarchist." *Philosophical Studies* 81 (2–3), 247–61. https://doi.org/10.1007/BF00372785

Schuster, Radek. 2018. "The Analogy between Goodman's Riddle of Induction and Wittgenstein's Paradox of Rule-Following." In *Philosophy of Logic and Mathematics*, edited by Gabriele M. Mras, Paul Weingartner, and Bernhard Ritter, 224–26. Kirchberg am Wechsel: ALWS.

Organon F 27 (3) 2020: 325–345
https://doi.org/10.31577/orgf.2020.27303
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Physical Constants as Identifiers of Modern Universal Laws of Nature

Orli Dahan*

Received: 18 January 2019 / Accepted: 3 December 2019

Abstract: I argue that in modern algebraic-formulated science the 'physical constant' can be understood, for practical purposes, as an 'identifier' of a universal law of nature. This identifying role is possible because the concept of 'physical constant' fulfills the same need for universality, stability, and fundamentality (as universal laws) for increasing the epistemic value of a scientific theory. This can be demonstrated in two different ways. The first involves a thought experiment envisioning science without physical constants, which appears to be a science of local and particular laws. The second is the observation that physical constants mostly emerge as components in an algebraic formulation of universal laws, but not in the algebraic formulation of particular laws. This observation about the link between physical constants and universal laws of nature, if correct, makes two contributions. First, it clarifies, at least partially, the ambiguity in the use (and the absence) of the concept 'law' in contemporary science. Second, it can help in distinguishing between a universal law and a particular law, while avoiding one of the abiding philosophical problems regarding laws of nature—the problem of the *ceteris-paribus* criterion for a generalization.

Keywords: Law of nature; universal law; particular law; physical constant; *ceteris-paribus* generalization; *strict* generalizations.

* Tel-Hai College
Faculty of Humanities and Social Sciences, Tel-Hai College, Upper Galilee, 12208, Israel
orlydah@telhai.ac.il

1. Introduction

Physical constants are essential to current physical theories; they are values in many physics equations. Such constants include the velocity of light (c), Newton's constant of gravitation (G), and Planck's constant (h).[1] Physical constants are dimensional quantities with two fixed parameters: they are both constant at every point in space and have a constant value in time. In fact, physical constants have become so important to modern science, that it seems that every new grand physical theory introduces a new constant. For example, string theory and its coupling constant (Greene 2003, 303–06), or GRW in quantum mechanics and its new collapse constant (Frigg 2009).

In this paper, I argue for a conceptual link between modern physical constants and the concept of universal laws of nature. Defining 'law of nature' ('what is it to be a law?') is not an easy task. Definitions vary widely among scientists and philosophers of science and are dependent upon one's metaphysical commitments (see Caroll 2016). However, choosing a side in this age-old debate will not contribute to the primary goal here. So, for the purpose of this paper, I choose to define superficially the term 'universal law' via negation—contrasting it with the term 'particular law.'[2] By 'universal law,' I mean a general statement that governs a class of phenomena and is *universal* in the sense that it has no exceptions: it appears to be applied everywhere in the universe in the same way. In this sense it is *absolute* and *stable*. A straightforward example of such a law is the law of inertia. It is universal in the sense that it governs all bodies in the world. In contrast, by 'particular law' I mean a general statement that governs a class of phenomena, but is *particular* in the sense that it cannot be applied

[1] About 20 physical constants are known today, including, also, the rest mass of an electron (m_e), Avogadro constants (N_A), and Boltzmann's constant (k_B).

[2] Definition via negation is popular in the field of philosophy of mind, where there is a long-standing discussion regarding how to define the concept 'physical' (see Crane and Mellor 1990). The most accepted way currently to solve this dilemma is 'via negativa.' This solution proposes to render the term 'mental' as fundamental and to characterize the physical as 'non-mental,' i.e. defining the term 'physical' negatively (Dahan 2019; Fiorese 2016; Prelević 2017).

everywhere in the universe. In this sense it is *conditional* (rather than absolute). A good example is natural selection. Natural selection is a key mechanism of evolution (the change in the heritable traits characteristic of a population over generations), but it is not universal, for it is applicable only to organisms and not to non-vital objects. Another example from physics itself is Ohm's law, which is conditional, for is not applicable in extreme temperatures, or in some electrical components such as semiconductors.[3]

Caroll (2016) considers whether philosophy can advance beyond the dispute over laws of nature given the question of whether lawhood is a necessary component of scientific theories. We can see that in current physics, the status of the nomic term "law of nature" has practically fallen from grace, as opposed to (for example) its status at the time of the scientific revolution (see Ruby 1986; Shapin 1996; Zilsel 1942).[4] For example, we still use the term "the laws of motion," or "Newton's second law," but we do not refer to the equations of quantum mechanics (QM) as 'laws.' Are the uncertainty principle or Schrödinger equation laws of QM? Are they universal laws in the meaning that "Newton's second law" was considered a universal law? According to Reutlinger et al. (2017), the Schrödinger equation clearly fits the traditional concept of a universal law. Either way, it can be said that the term "law" is absent from current, formal statements of modern theoretical physics. Indeed, according to Milton (1981), the discoveries of modern physics since Maxwell (1831–1879) have for the most part been expressed in terms of principles and equations rather than laws.

[3] This particular example will be discussed in Section 3.

[4] According to Shapin (1996), sixteenth and seventeenth century scientists and philosophers disputed the question whether science must be a mathematically formulated binding set of laws of nature, or whether mathematical representations could actually capture the contingencies and complexities of real natural processes (Shapin 1996, 58-59). Robert Boyle (1627–1691), for example, was uncomfortable with the common understanding of the concept 'laws of nature,' and repeatedly cautioned that it should be used carefully—moving from observed regularities to laws of nature obviates the factor of God's pleasure, power, and willing (Shapin 1996, 150). This discussion is evidence that the concept 'law of nature' was central at that time.

I carefully assume that maybe the partial absence of 'laws' in theoretical physics is due to the existence in modern physics of other underlying principles, such as symmetries and super-symmetries, which are not laws but rather *constraints* upon phenomena, laws, and theories.[5] Such a constraint can be seen in Fermat's principle. This principle is not a law, but serves as a guiding principle in the formulation of physical laws using the calculus of variations.[6] Other constraints are symmetry principles in physics, such as Noether's theorem, which states that every differentiable symmetry of the action of a physical system has a corresponding conservation law (Rosen 1989).[7]

Nonetheless, even in current science we *do* use the term 'law.' However, this term, it seems, is used in science without distinction between universal law and particular law. This also happens in the field of physics itself. For example, we use the term 'law' in order to refer to "the law of gravitation," which is considered a universal law, and we use the term 'law' to refer to "Ohm's law," which is not considered a universal law. The same can be said with regard to the "ideal gas law" (universal law) and the "Boyle-Mariotte law," a special case of the ideal gas law, hence not a universal law. As noted, the ambiguity is also in the other direction: laws in fundamental physics are not called 'law' anymore. For example, the Schrödinger equation in QM is called an 'equation' even though it is one of the important laws of QM. This ambiguity in the use (and the non-use) of the term 'law' in modern physics calls for explanation and clarification, for it is possible that in

[5] The relationship between laws and constraints in physics is worth examination. For example, can a law be a specific type of constraint? This discussion, however, is beyond the scope of this paper, and the answer to this question will not affect the arguments presented here.

[6] For an empirical example see (Westphal et al. 2002).

[7] Rosen (1989), for example, shows that when following the definition of symmetry as invariance under transformation, both reproducibility and predictability are kinds of symmetry. And since reproducibility and predictability are the two most fundamental principles of science—he argues that symmetry not only serves *within* science but *actually lies at its very foundation.* In my opinion, Rosen's approach is as an example for the falling from grace of the term "law of nature" in the realm of theoretical physics during the 20th and 21st centuries, and the rise of new terms (other than 'universal law of nature') that are considered fundamental and more essential.

algebraic-formulated fundamental science we use the term 'law' vaguely, or sometimes do not use it at all.[8,9]

The straightforward conceptual link between the two concepts—'law of nature' and 'physical constant'—is obvious: both are connected to fundamentality, universality, and stability. However, it is also clear that these terms refer to different entities with different logical formulations—a necessary relation in the case of laws of nature, and a contingent quantity in the case of physical constants. Nevertheless, from the introduction it follows that there are (at least) two main reasons that a philosophical investigation of the link between physical constants and laws of nature is needed. The first reason is that the latter term is still in use, yet its use is typically ambiguous. The second reason is that a new concept appeared in the physics of the late 19th century—the 'physical constant'—which became an important component of physical theories,[10] and has a straightforward, conceptual link to the senior concept 'universal law of nature.'

In this paper, I propose a method for distinguishing between universal and particular 'law' using the concept of physical constants. I will suggest the observation that physical constants generally emerge as components in the algebraic formulation of universal law, but not in the algebraic formulation of particular laws. Thus, I will investigate the hypothesis that physical constants, although clearly referring to different entities, can be used

[8] In fact, it would be interesting to check whether all laws termed specifically as 'laws' in modern physics are older than the 20th century (before the age of physical constants) while newer 'laws' are not called laws anymore (such as in QM and relativity). If true, the question is then whether the source of this change is the ambiguity in the concept 'law,' the shift to algebraical formulations in physics, or another reason. However, analysis of the historical reasons for the use and non-use of the term 'law' in modern physics is beyond the scope of this paper.

[9] A note is needed here: I am primarily making a sociological-historical claim about how scientists talk about 'laws,' 'constraints,' and 'constants,' and not a philosophical claim about how I think scientists should talk about these terms.

[10] Newton's constant of gravitation (G) was clearly the first universal physical constant to appear, but its value was empirically calculated only at the end of the 18th century by Henry Cavendish (1731–1810; Clotfelter 1987). Furthermore, the constant was given a symbol (the letter G) only in 1873, 186 years after Newton's *Principia* (in Cornu and Baille 1873).

for testing the way we think of a certain phenomenon, field, branch of science, or even science in general: whether a phenomenon is fundamental in the sense that it is governed by universal laws, or whether it is (only) local and governed by particular laws. Using a thought experiment, I suggest that in modern science, the role of physical constants differs in theories that have differing scopes.

In Section 2, I present my thought experiment: what would physical science look like if we came to understand that the fundamental physical constants are not constants at all and, in fact, that there are no such stable physical quantities. The thought experiment is a framework for thinking about one of the functions of physical constants in science, and for suggesting my reflections on the conceptual link between constants and universal laws. In Section 3, I give examples to further the following objectives: (a) to support my hypothesis that physical constants emerge as components in the algebraic formulation of universal laws, but not in the algebraic formulation of particular laws; (b) to further clarify my philosophical principle that physical constants are identifiers of universal laws, and the limitations of this principle (i.e. only in algebraic formulated science); and (c) to demonstrate how my principle can help us distinguish between a universal law and a particular law, while avoiding one of the abiding philosophical problems regarding laws of nature (the *ceteris-paribus* criterion for a generalization). In Section 5, I conclude this philosophical investigation and in Section 6, I discuss possible implications for the ideas presented.

2. The thought experiment: modern physics without physical constants

To set the ground for my thought experiment, a short analysis of the concept "physical constant" is in place. What is it to be a physical constant? Or, stated differently: what are the necessary conditions for a physical quantity to gain the status of a physical constant? The straightforward answer is that this quantity must be unchangeable in space and in time, in the meaning of 'steady.' This simple definition also applies to one of the features of a universal law of nature, which is supposed to govern all phenomena at

all points in space, and at all times. In fact, this is, I think, the meaning of 'universality.' So, in this simple matter of superficial definitions, we can see that there might be a connection between the two concepts.

Now, suppose we were to discover in the near future that all fundamental physical constants are not constant at all. More precisely, we learn that we were wrong to believe that these physical quantities are fixed quantities: either we come to understand that the constants are not constants (for example we will find out that the speed of light in a vacuum is not the maximum speed of massless particles), or that they do not exist at all (something like the mistake Einstein made with the cosmological constant).[11] The question this thought experiment is addressing, which I will answer throughout this paper, is the following: how does this imaginary scenario affect physics?

To answer this question, we first need to ask what makes a physical constant, constant, and how do we come to understand that a certain physical quantity is not really a constant physical quantity?

In fact, the thought experiment I propose is not so far-fetched, for there are various discussions in the literature regarding this possibility.[12] Moreover, the reverse of this thought experiment is basically how we *give* a specific dimensional physical quantity its status as a physical constant, for in some cases, we come to understand that a certain dimensional physical quantity is universally constant after initially thinking it a 'particular' constant. The most straightforward example is the speed of light c: when it was first measured, it was only a characteristic of a particular system.[13] However, with the development of classical electromagnetism, it became a characteristic of a class of physical phenomena (electromagnetic phenomena). Only with the

[11] For more about Einstein's mistake regarding the cosmological constant (Λ) see (Harvey and Schucking 2000). For a review of the basic physics and astronomy of the subject and the history of this idea see (Peebles and Ratra 2003).

[12] For example, Mangano et al. (2015) consider the possibility that the Planck constant is a time dependent quantity, undergoing random Gaussian fluctuations around its measured constant mean value. Support for this possibility is derived from Dirac's idea that fundamental constants are dynamic variables and from conjectures on quantum structure of space–time at small distances.

[13] For the history of measuring the speed of light see (Romer and Cohen 1940).

discovery of special relativity did the speed of light become a physical constant (Lévy-Leblond 1979).

As noted above, physical constants have two fixed parameters: they are dimensional quantities that are both constant at every point in space and have a constant value in time. But are these two parameters necessary for the constants to be *universal* in nature? According to Uzan (2011), "Any [physical] constant varying in space and/or time would reflect the existence of an almost massless field that couples to matter. This will induce a violation of the universality of free fall. Thus, it is of utmost importance for our understanding of gravity and of the domain of validity of general relativity to test for their constancy."

In principle, however, it is conceivable to have a physical constant that changes in time, but stays fixed for all points in space. For example, if we were to discover that G, the universal gravitational constant, changes slightly in time (periodically or linearly), and this change in time has some fixed regularity that is, in principle, determinate, quantifiable and predictable—then G is still a physical constant because it is constant at any point of the universe at a certain time—it is not local.

However, if we were to discover that G's value on earth is not the same as on one of Jupiter's moons, then it appears that we will have to deprive G of its status as a physical constant. And why? Because it is obvious that what is really unique about the physical constants is that they are constant at every point in space *at a certain time.* Now, if we discover that G is not a constant, the true implication of this discovery will be that the law of gravitation is *not universal,* as we thought it was. In this case, we will have to accept the fact that this law is a *local law,* for it does not govern the phenomenon of free fall on one of Jupiter's moons in the same manner it governs this phenomenon on earth.

In order to examine the implications for science of the elimination of physical constants, must I check all (around) 20 physical constants known today? I believe the answer is negative and that it is sufficient for the purpose of this thought experiment to consider only the three fundamental constants, G, c and h.[14]

[14] In fact, Matsas et al. (2007) argue that even these three can be reduced to just two fundamental constants (they prefer eliminating G, but apparently any of the

Let us now consider the other two fundamental physical quantities: the speed of light (c) and the Planck constant (h). As noted above, initially the speed of light was not considered a universal constant, so it is simple to imagine how reversion to this understanding will affect current physics. If, in the near future, we come to realize that c is not a constant quantity,[15] this 'constant' will probably revert to its former status as a characteristic of a particular system, or a characteristic of electromagnetic phenomena. But how will that affect what we think of electromagnetic phenomena? Will it still be considered a fundamental phenomenon? I hypothesize that if we discover that the other constants relevant to electromagnetic phenomena, such as the Planck constant (h) and electrostatic constant (k_e), are not constants, then the answer is affirmative: we will probably consider that electromagnetic phenomena are not universal. They may be considered local phenomena, or perhaps candidates for a special case of more fundamental phenomena. Consequently, the laws of electromagnetics will be considered particular laws and not universal laws. And what would that do to relativity theory? If c is no longer a physical quantity independent of any frame of reference—then it seems that relativity theory can no longer be considered a fundamental theory. Relativity theory cannot be a theory describing the whole universe, for it is not universal anymore.

The same point can be made regarding the Planck constant h: this constant is the most central to quantum mechanics. If we come to understand that this constant is not fixed at all places in the universe—it seems that we will have to construct a completely new fundamental physics.[16]

---

three is a candidate), and that their proof is model independent. This also indicates that the long search of fundamental science for a universal law of nature (an underlying law that governs all there is) has shifted in contemporary science. In physics today, 'laws' are just another component of a scientific theory, and it even seems that the main focus of reduction is not laws, as it once seemed to be.

[15] For example, we will find out that c is not the finite speed of massless particles on Jupiter, and that it seems that only on Jupiter and beyond, the finite speed is 350,000 km/s.

[16] No one really knows how this 'new physics' will look, if we come to understand that the constants are not constants. As Barrow and Webb (2005) conclude: "The constants are a tantalizing mystery. Every equation of physics is filled with them, and they seem so prosaic that people tend to forget how unaccountable their values

For example, Lizzi et al. (2016) discuss the possibility that the fundamental Planck constant has stochastic fluctuations, set by a possible quantum structure of space-time. It is worth noting that although being time dependent is a nontrivial affair for fundamental dimensional quantities, this possibility is being discussed currently in the field of theoretical physics. However, I have not found discussions in the literature of physical science regarding the possibility that this constant (or the other fundamental constants) may not be constant in space. This possibility would change the quantity from universal to local and undeserving of the status of 'physical constant;' yet, determinable regular changes in time while remaining constant in space would not deprive h of its characterization as a physical constant.

Since modern day universal laws of physics *are* essentially these universal equations, if we come to understand that there are no true physical constants in fundamental science, it seems that *we will have to give up universality within the equations themselves: we will have only local laws.* This will not necessarily affect other components of physical theories (such as underlying principles). However, laws that are expressed in modern science in the form of algebraic equations will have to be regarded as local laws only.

This thought experiment demonstrates that what is unique about physical constants is that they are fixed for all points in space. If we waive that requirement—then what we get is *locality*, which is (in this context) the opposite of *universality.* Thus, it seems to me that current universal laws whose physical constants are not strictly universal (the value can remain in the equation as a sort of 'local quantity;' as with c before general relativity) actually become particular laws. Hence, I draw the following hypothesis:

i. Algebraic science (like ours) without physical constants is the science of particular laws that govern (in the best case) only a class of physical phenomena.

---

are. Their origin is bound up with some of the grandest questions of modern science, from the unification of physics to the expansion of the universe. They may be the superficial shadow of a structure larger and more complex than the three-dimensional universe we witness around us. Determining whether constants are truly constant is only the first step on a path that leads to a deeper and wider appreciation of that ultimate vista."

From this it (carefully) follows that:

ii. A law constructed in an algebraic modern formulation that is based upon a physical constant *prima facia* fits the definition of a 'universal law.'

Hence:

iii. In an algebraic formulated science (like ours), a physical constant *can be seen,* at least in practice, as an *identifier* of a 'universal law of nature.'

This conclusion can also be described as a philosophical principle:

iv. In an algebraic formulated science, the presence or absence of physical constants can be used in order to distinguish between universal laws and particular laws respectively, in the manner I suggested in the beginning of the paper. Put more generally, physical constants can be used as an identifier for testing the way we think of a certain phenomenon, field, or branch of science; i.e. whether a phenomenon is fundamental and universal, or whether it is local and particular.

Two qualifications are required:

The first note is that this is *not* a knock down argument, but rather a pragmatic argument for those trying to better understand and clarify current scientific practice. It is based on an observation regarding the appearance and absence of physical constants in physical equations, and a semantic analysis of the practical meaning of what it is to be a 'physical constant' in relation to a 'universal law.' For this reason, the principle I suggest is to be used carefully. Nonetheless, in Sections 3, I provide examples that support this principle.

The second note relates to the condition "in an algebraic formulated science." This condition is meant to emphasize that I am not arguing that physical constants are metaphysically necessary for formulating a universal law of nature.[17] This is due to the fact that physical constants emerged in

[17] Many discussions in current literature concern the question of why physical constants have the values they have, and what would happen if they had other values (e.g. anthropic principles), assuming, perhaps, that physical constants reveal some

science only after the development of calculus and, in fact, it seems that the physical constants emerge when we translate a universal law (described verbally or geometrically) into an algebraic equation. Various examples demonstrate this emergence, such as the electrostatic constant in Coulomb's law, or the gravitational constant in Newtonian theory—the later will be discussed in Section 3. When a law is not expressed algebraically, there is no physical constant involved.[18] This does not mean that the law is not universal.[19] For example, it is in fact possible to imagine hypothetical intelligent creatures constructing an advanced science using a completely different language (a new geometry or other unknown language) that is capable of describing and explaining all the phenomena that our algebraic formulated science currently describes and explains—without the need for a single physical constant.[20]

---

deeper truth about order in the universe, or the meaning of the universe (see Smeenk and Ellis 2017). This paper, however, examines the concept of physical constants from a metaphysical *neutral* point of view, for it is possible that constants are simply a tool of a specific language. Even if they are only a tool of calculus, the examination is of interest given the relation of physical constants with the concept of 'law of nature.' The kind of discussion I engage in here is, as far as I know, absent from current literature.

[18] In Coulomb's original book (Coulomb 1785), it is clear that like in the *Principia,* Coulomb's law is introduced in a *descriptive manner,* without a physical constant. As with the gravitational constant (*G)*, the electrostatic constant (k_e) appears only later, when the law is formulated algebraically.

[19] The law of inertia can also be given as an example of a universal law that does not have a constant. However, it is not formulated algebraically, so my principle does not apply.

[20] There is a view that distinguishes between dimensional and dimensionless constants. For example, according to Duff (2015), dimensional constants, such as *h, c, G, e,* or *k*, are merely human constructs whose number and values differ from one choice of units to the next. On the other hand, only dimensionless numbers such as the fine structure constant, α, are independent of one's choice of units or measuring apparatus. Duff argues that in this sense only dimensionless constants are fundamental. However, the distinction and principle I suggest do not, in any sense, take a side in this issue. My purpose is different, and although related to the question of fundamental (or universal) laws—does not imply anything regarding the necessity nature of physical constants.

3. Test cases and limitations: physical constants as identifiers of universal laws

In this section I introduce and briefly discuss two test cases in order to contextualize and further clarify my argument. The discussions presented have three purposes. The first purpose is to support my hypothesis that physical constants emerge as components in algebraic formulations of universal laws, but not in algebraic formulations of particular laws. The second purpose is to further clarify my philosophical principle that physical constants can be seen as identifiers of universal laws, and the limitations of this principle (it can be applied only in algebraic formulated science). The third is to show that in a way, my principle can help distinguish between a universal law and a particular law, while avoiding one of the long standing philosophical problems regarding laws of nature—the *ceteris-paribus* criterion for a generalization.

3.1. Kepler's laws versus Newton's laws[21]

Kepler's laws (Kepler [1609] 1937; [1619] 1997) were intended to describe only the solar system, whereas Newton's laws were supposed to be universal

[21] It might appear that using Newtonian gravitation is not exactly the best example to a universal law, because Newton's law of gravity does not apply at speeds approaching the speed of light or in high gravitational fields, so how is it universal? A clarification is needed: It is, in fact, possible to imagine a hypothetical possible universe, in which the true physical theory is the Newtonian physics. It is clear that in that imaginary universe, Newton's law of gravitation is a truly universal law. But the truth value of a theory has nothing to do with whether it is universal or particular. In our universe, we currently refer to general relativity theory as the theory that governs the universe—but we cannot be sure that this theory is the true theory. The "true" theory of our actual universe could turn out in the future to be something completely different. However, whatever will become of fundamental physics in the future—the form of general relativity, exactly as the form of Newtonian physics, is general and universal, whether it is true or not. In other words, whether we identify a law or a theory as universal or not, has nothing to do with its truth value, but it depends on its pretense: if it was true—it is supposed to be applied to any place in the universe. Furthermore, in the case of Newton's law of gravitation—it is not a true law. It so happens that we use it, for our convenience, because in most cases

(Smith 2008). Isaac Newton (1642–1727) devised his laws of motion and gravitation (Newton 1687) in such a way that Kepler's laws can be derived from them (Smith 2008).[22]

That Kepler's laws are not universal is not entirely straightforward, in a sense, because Kepler's laws are not only true for the solar system, but for any system of planets that have a gravitational force that is proportional to the inverse square of the distance between them. For example, they are true for systems of binary stars, or systems of a moon and a planet. But arguing that Kepler's laws are universal laws on this basis begs the question, for it is equal to saying that Kepler's laws are valid for every system to which the laws apply. This obviously empties the concept 'universal' of its meaning, and it is also not the universality for which we are looking.

Kepler's laws are an example of one of the long-standing disputes about laws of nature: the problem of the distinction between *strict* generalizations (that are usually taken to be, at the very least, true, universal statements that support counterfactual claims) and *ceteris-paribus* generalizations (that are usually taken to "have exceptions," to be "non-universal" or "to be *ceteris paribus* laws") (Reutlinger et al. 2017). The claim is that while in theory this distinction is easy enough to understand, in practice it is often difficult to distinguish strict from ceteris-paribus generalizations, because many statements with no explicit ceteris-paribus clause implicitly do have such a clause (Carroll 2016). In our example, the ceteris-paribus clause in Kepler's laws is explicit—for it is not true for systems that are not of

we are not approaching the speed of light or are found in high gravitational fields. Physicists say that the Newtonian theory is approximate at low speeds. But the term 'approximate' implies that it is not accurate—only close enough. It doesn't change the fact that the Newtonian laws do not govern our world, and if we want to be 100% accurate, we need to use other theory, and not Newton's theory.

[22] Johannes Kepler (1571–1630) built upon Copernicus's work to create a much more accurate description of the solar system (Rabin 2015). The first law establishes that the orbit of a planet is an ellipse with the sun as one of the foci. According to the second law, the radius vector from the sun to a planet sweeps out equal areas in equal times, which means that the planet travels faster when closer to the sun and slower when farther from the sun. The third law (the "harmonic law") captures the relationship between the distance of planets from the sun and their orbital periods (Di Liscia 2017).

binary stars. However, one can argue that the ceteris-paribus clause also exists in Newton's law of gravitation. For example, Cartwright (2002) is usually understood to be an advocate of ceteris paribus laws in science, including physics. More precisely, Cartwright is understood to be *against* the use of laws in scientific explanations, including physics, exactly because all laws in physics use ceteris paribus. Indeed, all areas of science use idealistic assumptions to simplify things: physicists describe motion on frictionless surfaces because it is too difficult to explore everything at once. A simplistic model removes confusing factors and focuses on a specific parameter. Furthermore, when it comes to testing, we test *ceteris paribus* laws in exactly the same way that we test laws without the *ceteris paribus* antecedent. However, the presence of a *ceteris paribus* antecedent forces us to take into account important procedures when designing experiments.

This can be somewhat confusing, since knowing that the laws of Kepler can be derived from Newton's laws makes us think, with considerable confidence, that Newton's laws are universal, while Kepler's laws are particular. Minimally, we might think that Newtonian laws are *more* universal than Kepler's laws. Hence, the distinction between *strict* generalizations and *ceteris-paribus* generalizations is insufficient in practice.

When converted to algebraic equations, in both Kepler's laws and Newton's laws a constant emerges.[23] However, the fixed ratio in Kepler's third law does not remain steady but is different for each system. This constant depends upon the mass of the planets in the system, so when the masses are different—the constant changes. In contrast, in Newton's law of gravitation, the constant is universal, for it remains steady across space and types of physical systems no matter what their mass. This explains why the constant in Kepler's law is not a physical constant in the same sense I am dealing with in this paper. Hence, it is clear that the principle I am offering

[23] This outcome demonstrates that a constant is in no way a necessary element of a universal law of nature—it is, in fact, a consequence of the translation of a universal law to an algebraic equation. For the sake of discussion, if we were constructing a fundamental science in some other language (such as verbal or geometric), perhaps there would be no need for physical constants at all in order to describe a universal law of nature.

helps us distinguish between a universal law and a particular law while avoiding the problem of the *ceteris-paribus* criterion for a generalization.

3.2. One more particular law without a constant

Ohm's law (Ohm 1905, original work published in German 1827) is an equation ($R = V/I$) that describes the relationship between the current through a conductor between two points in units of amperes (I), the voltage measured across the two points in units of volts (V), and the resistance of the conductor in units of ohms (R). This law was named after the German physicist Georg Ohm (1789–1854), who carried out and described measurements of applied voltage and current through simple electrical circuits containing various lengths of wire. Ohm's law is an empirical law, that is, a generalization from many experiments that have shown that the current is approximately proportional to electric field for most materials.

According to Weber et al. (2012), in the early 20th century, it was thought that Ohm's law would fail at the atomic scale, but in fact it was observed on a wide range of scale lengths. Moreover, Weber et al. (2012) demonstrated that Ohm's law works for silicon wires as small as four atoms wide and one atom high. However, Ohm's law is not considered a universal law, because any given material will break down (electrical breakdown) under a strong-enough electric field, and some materials of interest in electrical engineering are "non-ohmic" under weak fields (Griffiths 1999, 289). Thus, Ohm's law is a particular law of a certain class of phenomena, because it cannot show a linear, determinate and regular relation between voltage and current maintained across time. For example, in extreme temperatures the linear ratio is not maintained, and there are electrical components, such as semiconductors, for which the ratio between voltage and current is a non-linear and unpredictable one.

I believe that this example strengthen the philosophical principle I have suggested regarding physical constants as identifiers of universal laws. Similar to the case of Kepler's laws, Ohm's law's particular status seems to follow from a non-universality of matter. Moreover, there is no physical constant in Ohm's law formulated algebraically.

4. Concluding remarks and further implications of the idea

I have argued that at least in current science, there is a link between physical constants and universal laws of nature. I suggested that a physical constant can be seen as *an identifier* of a universal law of nature in algebraically formulated science. This observation might help us avoid problems in defining a universal versus particular law, such as *strict* generalizations versus *ceteris paribus generalizations*. It can also help clear ambiguity in the use of the term 'law' in modern physics, for it seems that we use the term 'law' rather loosely.

I started with two basic observations. The first was that in modern physics, the nomic term "law of nature" (or 'universal law of nature') has practically fallen from grace. While this term is still in use, we use it without differentiating between particular and universal law. Moreover, there are many cases in fundamental physics where universal laws are not called 'laws' at all, but 'equations' or 'principles.' The second observation is that a new term emerged in the late 19th century and early 20th century—the 'physical constant.'[24] The physical constant became so important in current, fundamental science that every new theory 'baptizes' a new physical constant. On the face of it, I have indicated that there is a superficial link between these two concepts: both laws of nature and physical constants are supposed to be universal and fundamental.

In order to test this link, I constructed a thought experiment, asking the question how science would look void of constants, for example, if we were to find that the three fundamental constants c, h and G are not fixed at every point in space. The answer was that it will be a science of particular laws, governing classes of phenomena. If my philosophical principle is correct, it can be used as a method to distinguish between a universal law and a particular law. Thus, physical constants can help us test the way we think of a certain phenomenon, field, or branch of science—whether a phenomenon

[24] One can emphasize the technical aspect of my observation: clearly the use of physical constants emerged in science due to the use of the algebraic language of calculus and the rise of the international system of units (SI), thus there is nothing special about these quantities. However, I am not arguing here that these physical quantities are a metaphysical necessity.

is fundamental and universal, or whether it is (only) local and particular. For example, I suggested that there is a sense in which both the 'uncertainty principle' and Schrödinger equation, though referred to in physical science as 'principle' and 'equation' respectively—in fact deserve the title 'universal law.' Nevertheless, other so called 'laws' in science are not universal laws (such as Ohm's law or Kepler's laws) but particular generalizations.

I believe that my hypothesis offers new ways to investigate these notions from a philosophical point of view. According to Carroll, "more attention needs to be paid to the language used to report what are the laws and the language used to express the laws themselves. It is clear that recent disputes about generalizations in physics and the special sciences turn on precisely these matters, but exploring them may also pay dividends on central matters regarding ontology, realism vs. antirealism, and supervenience" (Carroll 2016). In a way, this paper is doing exactly this: paying attention to the contemporary language of universal laws, which is, in our case algebraic.

A further implication of the proposition that a physical constant can be an identifier of a universal law of nature is the following hypothesis: since the rise of physical constants, they can be seen (de facto) in an algebraic formulated science as a *substitute* for the historical concept 'law of nature.' This might answer Caroll's (2016) question how can philosophy advance beyond the long standing disputes over laws of nature. An interesting possibility is that perhaps lawhood is not supposed to be part of the content of *algebraic* formulated scientific theories. In that formulation we have other, more suitable content in the form of a physical constant, which fulfills the same need for universality, stability, and fundamentality (as universal laws)—but from *within* the theory itself. This claim of 'replacement' is surely stronger than what I suggested here regarding the physical constants as identifiers of universal laws. If taken seriously, the claim must be developed further, because the only connection between universal laws and physical constants may be that both a universal law and a constant are universal, and that both are invariant across time and space. That connection is clearly insufficient for a stronger claim of replacement.

Lastly, the principle I suggested might also shed light upon an old controversy: can there be any special-science universal laws? If my hypothesis is correct, then the answer is basically affirmative—if a universal and

invariant constant is rooted in the equations of a special science, its presence would imply that the law is universal and not particular.

Acknowledgements

I am grateful to Giora Hon, Yael Kedar, and Ayelet Shavit, for providing feedback on an earlier draft of this article. I thank Joseph Agassi for a helpful conversation concerning physical constants, and to Gad Freudenthal, and Raz Chen-Morris for a useful discussion about laws of nature.

References

Barrow, John D., and Webb, John K. 2005. "Inconstant constants." *Scientific American* 292 (6): 56–63. https://doi.org/10.1038/scientificamerican0206-64sp

Carroll, John W. 2016. "Laws of Nature." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2016/entries/laws-of-nature/

Cartwright, N. 2002. "In Favor of Laws that Are Not *Ceteris Paribus* After All." In *Ceterus paribus laws*, edited by John Earman, Clark Glymour, and Sandra Mitchell, 149–63. Dordrecht: Springer. https://doi.org/10.1007/978-94-017-1009-1_8

Clotfelter, Beryl E. 1987. "The Cavendish Experiment as Cavendish Knew It." *American Journal of Physics* 55 (3): 210–13. https://doi.org/10.1119/1.15214

Cornu, A., and Baille, J. B. 1873. "Mutual Determination of the Constant of Attraction and the Mean Density of the Earth." *C. R. Acad. Sci., Paris* 76: 954–58.

Crane, Tim, and Mellor, D. H. 1990. "There is No Question of Physicalism." *Mind* 99 (394): 185–206. https://doi.org/10.1093/mind/XCIX.394.185

Dahan, Orli. 2019. "There IS a question of Physicalism." *Organon F* 26 (4): 542–71. https://doi.org/10.31577/orgf.2019.26401

Di Liscia, Daniel A. 2017. "Johannes Kepler." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2017/entries/kepler/

Duff, Michael J. 2015. "How Fundamental Are Fundamental Cconstants?" *Contemporary Physics* 56 (1): 35–47. https://doi.org/10.1080/00107514.2014.980093

Fiorese, Raphaël. 2016. "Stoljar's Dilemma and Three Conceptions of the Physical: A Defence of the Via Negativa." *Erkenntnis* 81 (2): 201–29. https://doi.org/10.1007/s10670-015-9735-0

Frigg, Roman. 2009. "GRW Theory (Ghirardi, Rimini, Weber model of quantum mechanics)." In *Compendium of Quantum Physics: Concepts, Experiments, History and Philosophy*, edited by Daniel Greenberger, Brigitte Falkenburg, Klaus Hentschel, and Friedel Weinert, 266–70. Heidelberg and Berlin: Springer. https://doi.org/10.1007/978-3-540-70626-7_81

Greene, Brian. 2003. *The Elegant Universe: Superstrings, Hidden Dimensions and the Quest for the Ultimate Theory*. NY: W.W. Norton & Company.

Griffiths, David J. 1999. *Introduction to Electrodynamics* (3rd ed.). Prentice Hall.

Harvey, Alex, and Schucking, Engelbert. 2000. "Einstein's Mistake and the Cosmological Constant." *American Journal of Physics* 68 (8): 723–27. https://doi.org/10.1119/1.19534

Kepler, J. [1609]. 1937. *Astronomia Nova*. Pragae.

Kepler, J. [1619]. 1997. *The Harmony of the World* (Vol. 209). American Philosophical Society.

Lévy-Leblond, J. M. 1979. "The Importance of Being (a) Constant." In *Problems in the Foundations of Physics, Proceedings of the International School of Physics 'Enrico Fermi' Course LXXII, Varenna, Italy, July 25 – August 6, 1977*, edited by G. Toraldo di Francia, 237–63. Amsterdam, New York: North-Holland.

Lizzi, Fedele, Mangano, Gianpiero, and Porzio, Alberto. 2016. "Planck's Inconstant." In *Proceedings of the Corfu Summer Institute 2015*. SISSA Medialab. https://doi.org/10.22323/1.263.0108

Mangano, Gianpiero, Lizzi, Fedele, and Porzio, Alberto. 2015. "Inconstant Planck's Constant." *International Journal of Modern Physics A* 30 (34): 1550209. https://doi.org/10.1142/S0217751X15502097

Matsas, George E., Pleitez, Vicente, Saa, Alberto, and Vanzella, Daniel A. 2007. "The Number of Dimensional Fundamental Constants." *arXiv preprint arXiv:0711.4276*.

Milton, John R. 1981. "The Origin and Development of the Concept of the 'Laws of Nature.'" *European Journal of Sociology/Archives Européennes de Sociologie* 22 (2): 173–95.

Newton, Isaac. 1687. *Philosophiae Naturalis Principia Mathematica*.

Ohm, Georg S. 1905. *The Galvanic Circuit Investigated Mathematically* (No. 102). D. Van Nostrand Company.

Peebles, P. J. E., and Ratra, Bharat. 2003. "The Cosmological Constant and Dark Energy." *Reviews of Modern Physics* 75 (2): 559. https://doi.org/10.1103/RevModPhys.75.559

Prelević, Duško. 2017. "Hempel's Dilemma and Research Programmes: Why Adding Stances Is Not a Boon." *Organon F* 24 (4): 487–510.

Rabin, Sheila. 2015. "Nicolaus Copernicus." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2015/entries/copernicus/

Reutlinger, Alexander, Schurz, Gerhard, and Hüttemann, Andreas. 2017. "*Ceteris Paribus* Laws." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2017/entries/ceteris-paribus/

Romer, M., and Cohen, I. Bernard. 1940. "Roemer and the First Determination of the Velocity of Light (1676)." *Isis* 31 (2): 327–79. https://doi.org/10.1086/347594

Rosen, Joe. 1989. "Symmetry at the Foundations of Science." *Computers & Mathematics with Applications* 17 (1–3): 13–15.

Ruby, Jane E. 1986. "The Origins of Scientific 'Law.'" *Journal of the History of Ideas* 47 (3): 341–59. https://doi.org/10.2307/2709657

Shapin, S. 1996. *The Scientific Revolution*. Chicago and London: University of Chicago Press.

Smeenk, Christopher, and Ellis, George. 2017. "Philosophy of Cosmology." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2017/entries/cosmology/

Smith, George. 2008. "Newton's *Philosophiae Naturalis Principia Mathematica*." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2008/entries/newton-principia/

Uzan Jean-Philippe. 2011. "Varying Constants, Gravitation and Cosmology." *Living Rev. Relativity* 14 (2): 1–155. https://doi.org/10.12942/lrr-2011-2

Weber, B., Mahapatra, S., Ryu, H., et al. 2012. "Ohm's Law Survives to the Atomic Scale." *Science* 335 (6064): 64–67. https://doi.org/10.1126/science.1214319

Westphal, Volker, Rollins, Andrew M., Radhakrishnan, Sunita, and Izatt, Joseph A. 2002. "Correction of Geometric and Eefractive Image Distortions in Optical Coherence Tomography Applying Fermat's Principle." *Optics Express* 10 (9): 397–404. https://doi.org/10.1364/OE.10.000397

Zilsel, Edgar. 1942. "The Genesis of the Concept of Physical Law." *The Philosophical Review* 51 (3): 245–79. https://doi.org/10.2307/2180906

Organon F 27 (3) 2020: 346–376
https://doi.org/10.31577/orgf.2020.27304
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

An Approach to Indexical Beliefs

Tadeusz Ciecierski*

Received: 11 June 2019 / Accepted: 3 February 2020

Abstract: An approach to indexical beliefs is presented and defended in the paper. The account is inspired by David Kaplan's representationalist analysis of *de re* belief reports. I argue that imposing additional constraints on the Kaplanian notion of representation results in an elegant theory of indexical beliefs. The theory is committed to representations of limited accessibility but is not committed to relativized proposition, special *de se* contents or propositions of limited accessibility.

Keywords: Belief *de hic*; belief *de nunc*; belief *de se*; indexical beliefs; propositions; representations.

The problem of indexical attitudes in general and *de se* attitudes in particular is the subject matter of vivid discussion in philosophy and linguistics. Some philosophers suggested even that addressing the problem properly requires the modification of the standard notion of truth-conditional content or demands a new theory of objects of attitudes. Without any doubt, the realization of any of the two expectations would have an enormous impact on numerous philosophical considerations regarding the nature of agents, actions, intentionality, and reference. This promise (for

* University of Warsaw
https://orcid.org/0000-0001-5403-9590
Institute of Philosophy, University of Warsaw, Krakowskie Przedmieście 3, 00-927 Warsaw, Poland
taci@uw.edu.pl

some) or threat (for others) suffices for the justification of the relevance of the problem in question. That having been said, however, it is difficult sometimes to say what the exact problem of indexical attitudes is as there are numerous issues discussed in the literature under this heading. It might be useful, therefore, to start with a short classification of the problems in questions. Following (Higginbotham 2003) we may start with the list of the following four questions:[1]

> **Problem 1**. What is the nature of *essentially indexical* interpretations of certain sentences or utterances?
>
> **Problem 2**. What relation do they bear to ordinary uses of the first-person pronoun, adverbs of time (like 'now') and place (like 'here')?
>
> **Problem 3**. Why are they triggered by the particular linguistic items that trigger them?
>
> **Problem 4**. Are they universal in human language, and what relation, if any, do they bear to logophoric phenomena?

and add to the list the following problem:

> **Problem 5**. What is the nature of essentially indexical attitudes?

Below I shall be interested solely in problems 1, 2 and 5 (by "the problem of indexical beliefs" I shall mean below nothing more than the three problems in question).

My general assumption here would be that the study of attitude reports or particular readings of such reports sheds light on the nature of corresponding attitudes. This assumption, I believe, is often made in the literature but rarely stated explicitly.[2] Consider any true attitude report of the form "A believes that p." It is natural to ask a question in virtue of which non-linguistic facts the report is true (or—to use fashionable terminology of truthamkers: which non-linguistic facts make the report true). The immediate answer to this question is that the report is true in virtue of the fact that A is in a certain *kind* of mental state. It is, of course, a matter of the

[1] Higginbotham's formulation of the problems concerns *de se* beliefs only but they might be easily generalized to other essentially indexical cases.

[2] For some exceptions, see (Vendler 1967; Peterson 1997; and Ciecierski 2016).

debate whether the state can be characterized solely internally or requires an appeal to external factors but this should not obscure the fact that the existence of the state in question is a necessary condition of the report being true. This general point is accepted even by eliminativists who deny not the connection between the existence of the state and truthfulness of the report but the fact that there are at all states of a certain kind (and hence: that the attitude reports have truth values at all). Now the natural next question is: "How the utterance selects, out of a vast universe of possible states of affairs, the ones that are potential truthmakers of the report?" It goes without saying that the answer must appeal to semantical properties of "A believes that p." The semantic properties, whatever they are (this, again, is a matter of vivid debates in philosophy), point out at *generic* features that each potential truthmaker of the report must be entitled with. This is how the assumption that the study of attitude reports reveals properties of the corresponding attitudes will be understood in this paper.

The purpose of this paper is to present and defend a relatively conservative approach to indexical beliefs and indexical belief reports. It shall be *conservative* as it requires no changes in our theories of truth-conditional content. It shall *not* be completely conservative as it denies that propositions are objects of attitudes. However, it shall still be *relatively conservative* as the direct objects of attitudes are familiar entities: mental representations. The plan of the paper looks as follows. I shall start with a brief recapitulation of the problem of indexical beliefs. Next, I shall describe an account in which I will offer a theoretical description of the phenomenon. Finally, I shall address several matters that, I think, may help us to better understand the scope and possible limits of the account.[3]

[3] Below I shall use terms such as "indexical belief" (Stalnaker 1981), "locating belief" (Perry 1979), "self-locating belief" (Lewis 1979; and Perry 2013) and "essentially indexical belief" (Stalnaker 1981 again) interchangeably. The terms such as "*de se* belief," "*de nunc* belief" and "*de hic* belief", on the other hand, will be used to refer to the proper subsets of the class of locating beliefs. I am not presupposing that the subsets are (or are not) disjoint.

1. Preliminaries

What are indexical beliefs? The standard answer to this question is that they are beliefs about "[…] where one is, when it is, and who one is" (Perry 1979, 5).[4] This answer is usually supplemented with scenarios that are intended to make the essentially indexical character of such beliefs explicit.

Consider a belief report:

(1) A believes that she is F.

This report might be true in two radically different situations: one in which A is aware of the fact that she herself is F and the other where she is unaware of that. The two situations might be more precisely referred to by means of the following paraphrases or readings of (1):[5]

(2) A believes that *she herself* is F.
(3) A believes that *she* (*but not that she herself*) is F.

Some suggest (cf. Jaszczolt 2013) that (2) and (3) are distinct interpretations of (1), and that this means that (1) is context dependent. Others may perhaps think that (1) is ambiguous. Yet others may be willing to treat (1) as a manifestation of the phenomenon of *generality*, that is the fact that

[4] A slightly different characteristic is provided in (Perry 2013): "A belief is self-locating if the truth of the belief constraints the location of the belief and the believer" (Perry 2013, 388). One may also explain the notion of indexical belief by saying that it is a belief state naturally expressed by 'I'-sentences, 'now'-sentences or 'here'-sentences (thanks to the anonymous reviewer for this suggestion). This intuitive description, however, is rather unclear as it adduces to the very imprecise concept of naturalness of expression.

[5] If one thinks that constructions like '(she) herself' do not univocally specify *de se* attributions (cf. Schlenker 2011), then the point is that there is a difference in truth conditions of counterparts of (2) and (3) that, respectively, contain or do not contain, elements that trigger *de se* attributions. Self-oriented long distance reflexives of Sells (1987) and Chierchia (1989) might be good candidates for such triggering elements. Let me just stress that nothing in this paper depends on the theoretical choice on that matter.

a meaning or a connotation of the expression is disjunctive.[6] The choice of the particular account here constitutes an answer to the Problem 3 and lies beyond the scope of this paper.[7] The common assumption of the alternative analyzes is that (2) and (3) differ in truth conditions.

This common assumption consists in the claim that (2) and (3) are true of situations that involve distinct *kinds* of belief states. This difference manifests in the dispositions and actions of the agent: if one accepts a functionalist picture of beliefs, one probably has to admit that the two belief states occupy different places in the functional web of beliefs, desires and actions. Philosophers believing in intentional explanations and predictions will notice that the regular intentional predictions in order to have any predictive and explanatory value must involve premises that have an indexical interpretation. Otherwise nothing interesting about agent's actions can be inferred.[8] This shows that indexical beliefs are ubiquitous as well as that the notion of indexical belief plays important role in practices of action explanation and prediction.[9]

---

[6] For a useful discussion about differences between *ambiguity*, *indeterminacy* and *generality* see (Gillon 1990).

[7] That being said, however, I would like to stress similarities between certain consequences of the present theory and the pragmatic theories that address the Problem 3. On the account defended in this paper, for instance, indexical attitudes are always attitudes *de re*. This analysis fits well the pragmatic theories (see Capone 2016) that make use of the indexical-*de re* entailment in order to explain the default character of certain (*de se*) readings in terms of scalar implicatures based on the *de se*-*de re* entailment scale.

[8] As John Perry once noted: when looking at our actions from the first-person perspective we quickly realize that: "When we replace $\langle\langle \mathrm{I} \rangle\rangle$ with other designations of me, we no longer have an explanation of my behavior and so, it seems, no longer attribution of the same belief" (Perry 1979, 3). This justifies, as Perry observes, classifying such beliefs as *essentially indexical*. As suggested by authors like Castañeda (Castañeda 1966) and Corazza (cf. Corazza 2004, 275–307) even a stronger claim holds: they are *essentially thusly indexical* as they involve a peculiar type of indexicality or mode of reference.

[9] It should be noted, however, that this broadly accepted belief has been questioned in recent years by authors like Cappelen and Dever (cf. Cappelen and Dever 2013).

It is also common to claim that indexical beliefs are philosophically interesting because they seem to put in question a popular picture of propositional intentional states. According to this popular picture (call it the *relationists view*) propositional attitudes are two-argument relations[10] that link cognitive subjects to abstract objects called "propositions." The relationists view assumes that propositions are truth evaluable entities that are eternally, impersonally and placelessly (and, in general, contextlessly) true or false. It embraces also the view that propositions are universally accessible.

What is the challenge posed by indexical beliefs? Consider the following scenario:

> *Absent-minded referee*
> Suppose that Monica received a paper to review. Monica is unaware that she is in fact its author (she might have submitted it a very long time ago and forgot about it). At a certain point she concludes that the paper should be rejected. Later she may (or she may not) realize that she in fact negatively evaluated her own work.

There can be no doubt that Monica's behavioral potential would have changed dramatically if she had acquired the indexical belief that she herself is the author of the paper. According to the view that combines relationism with anti-relativism the difference in the behavioral potential ought to be explained by the fact that in the two situations Monica is in the believing relation towards distinct propositions. Call the propositional content of indexical belief (whatever it is) an *indexical proposition*. Two options are prima facie available here: the indexical proposition is either general or singular.[11] However, both options are rather implausible. Monica might have

[10] Strictly speaking, relationism may enable more arguments in belief relation (e.g. time or place). The important point is that the additional arguments are not essentially linked to other representational properties of the cognitive agent.

[11] One might object to this description as committed to the false dilemma that ignores the possibility of singular contents of a special kind (special first-person *de re* senses, for instance). However, the commitment to such special *de re* senses conflicts with the assumption of universal accessibility. Hence such a theory is in conflict with relationism and anti-relativism.

already believed the general proposition that makes use of some descriptive concept D that uniquely refers to her. For instance, she may be the tallest woman ever born in Palo Alto and she might have been related to the proposition that she had rejected the paper authored by the tallest woman ever born in Palo Alto. Having such (or similar) belief does not preclude a huge change in her behavioral potential when she acquires the indexical belief. The only additional condition that has to be met is that she did not know earlier that she is D. The same seems to apply to singular propositions: she clearly might have believed that she rejected the paper authored by N (where N is a non-first person, directly referential term referring to her) before and still be the subject of the relevant belief expansion. Again, the only additional condition that has to be met is that she didn't know that she is N.

So, it seems that the combination of relationism and anti-relativism is unable to explain the relevant change in beliefs. Some authors (e.g. Feit and Capone 2013) call this issue "the psychological problem of *de se*." The challenge here is to "[…] identify the content of a given *de se* belief" and to distinguish it from contents of beliefs that are not *de se* (Feit and Capone 2013, 3). Additionally, they single out the linguistic problem of *de se* (what is the semantic value of "I," "s/he herself" or self-oriented long distance reflexives in belief clauses; the problem is a reformulation of Higginbotham's Problem 2) and the generality problem identified by Geach ("How distinct people can share the same *de se* belief?"). Geach's problem is closely connected with the problem of making sense of the intuitive assumption that distinct agents might perform the same type of action (Verdejo 2017). Indeed, if one dismisses the presupposition of Geach's question (namely that distinct people can share the same *de se* belief) it may be difficult to make sense of sameness of action that are—due to the involvement of *de se* attitudes—agent bound. In the section 3 below I shall discuss how the theory sketched below deals with the three problems.[12]

[12] In this paper I am not discussing problems that arise if a particular view about indexical beliefs is assumed at the starting point. For instance, if one accepts the Lewisian view that the proper analysis of indexical beliefs requires centered propositions, then the serious challenge is to reconcile this view with the standard models of linguistic communication (cf. Weber 2103; Rudnicki 2019).

2. Kaplan's account of *de re* belief reports

The approach to indexical beliefs I shall sketch below is inspired by the seminal analysis of *de re* belief reports by David Kaplan (Kaplan 1968).[13] Kaplan's theory attempts to address, in the first place, *the exportation problem*: the issue of providing general validity criteria for inferences that start with syntactically understood *de dicto* belief sentences as assumptions and arrive at syntactically understood *de re* belief sentences as conclusions.[14] For instance, everybody knows that the team that is going to beat its opponent in the next World Cup final is going to win the World Cup. This generally does not allow you to infer that there is a team of which you know that it will win the World Cup. It is clearly the latter knowledge, and definitely not the former, that may be of some interest to bookmakers. On the other hand, knowing that this team (assume that one points at a team having a training here and now) is going to win the World Cup entails that that there is a team *of which* you know that it is going to win the World Cup. So, it seems, in some cases believes-that or *de dicto* sentences enable us to derive believes-of or *de re* sentences.[15] The exportation problem is to make explicit the additional assumptions that warrant the inference.[16]

Kaplan's reply to the exportation problem, as I understand it, amounts to distinguishing two kinds of *de dicto* belief *states* and two kinds of corresponding belief reports. The first category embraces beliefs that involve a special kind of representation, the second category[17] does not. From the viewpoint of Kaplan's theory, we may say that *de re* beliefs are a special

[13] Kaplan himself abandoned this view in his latter writings (cf. Kaplan 1985, 326). Although my aim here is not to provide an exegesis of Kaplan's philosophical development I shall briefly address Kaplan's reasons of dissatisfaction in the section 4.3.

[14] Roughly speaking, a belief report is syntactically *de dicto*, if it contains a complete that-p clause, it is syntactically *de re*, if it contains the "believe(s) of [...] that it is F" predicate.

[15] We treat knowledge as a kind of belief.

[16] For a discussion of the exportation problem see (Kripke 2011).

[17] Here the distinction *de dicto*/*de re* applies primarily to states and secondary to attitude reports. As such it should be distinguished from the syntactical version of the distinction.

case of *de dicto* beliefs—the ones that make use of a specific class of representations. The crucial notion of representation is understood by Kaplan as a triadic relation ***R*** defined in the following manner (let α, β etc. range over vehicles of representation, whatever they are[18]):

R(α, *x*, *y*) (= α represents *x* for *y*) iff:
(i) α denotes *x,*
(ii) α is of *x,*
(iii) α is vivid for *y.*

I shall call the condition (i)—the *iconicity demand*, the condition (ii)—the *causality demand*[19] and the condition (iii)—the *vividness demand*.

I think that the best way of explaining the difference between (i) and (ii) is to consider the case through pictorial representations like photographs (Kaplan uses here the analogy with paintings). As Max Cresswell noted once:

> […] if one asks the question 'What is that a picture of?', one can be given an answer in terms of the causal ancestry of the picture. But perhaps the picture has been faked. […] There is still a legitimate answer to the question of what a picture is the picture of. This answer can be thought of as the answer to the question of what the world would have to be like in order for things to be as the picture shows them to be. A picture of a purple cow would demand at least a world in which there was a purple cow. (Cresswell 1985, 132)

Cresswell's point is that there are in fact two kinds of representational relations: one that is causally grounded (the photograph represents someone or something that has been photographed) and the other that is founded

[18] Since I do not want the described theory to be committed to any particular account of representations I am intentionally leaving the question about the nature of such vehicles open.

[19] A more adequate terminology might be that of *indexicality demand* as it is motivated by Peircean notion of indexical sign, that is the notion of sign that is existentially (in some cases: causally) connected with its object. I decided to talk about the causality demand in order to avoid the confusion with other uses of "indexicality" in this paper.

on the resemblance of the vehicle of representation and the thing represented (hence the *iconicity*). Kaplan's concept of representation presupposes the relevance of both dimensions: a vehicle of representation represents something (in Kaplan's sense) only if it resembles the thing represented and if the latter has played a special role in causing the vehicle's coming into existence. Since the relevant concepts of *resemblance* or *similarity*, according to Kaplan, apply not only to pictorial representations but also to linguistic expressions endowed with a descriptive content (like definite descriptions), that is the ones denoting objects in virtue of the fact that the objects have properties expressed in the description, resemblance must be understood here in a manner general enough to enable this. Roughly speaking, we may risk saying that Kaplan's general notion of denotation (and resemblance) amounts to presupposing that if certain facts about the denotatum had been different, the vehicle would not have been denoting this particular object. The "certain facts" in question are the ones in which the relation of generalized similarity is grounded. For instance, we might say that a road map would not have been denoting a particular road system if the number of crossroads had been different from the number of crossroad representations on the map. We might also say that, for instance, the description 'the inventor of topboots' would not have been denoting Wallenstein (the actual inventor of topboots) unless Wallenstein had had been the inventor of topboots.[20]

The concept of *vividness* concerns, as Kaplan puts it, purely internal aspects of individuation:

> The crucial feature of this notion is that it depends only on [...] current mental state, and ignores all links whether by resemblance or genesis with the actual world. (Kaplan 1968, 201)

[20] Recently Grabarczyk (Grabarczyk 2015) introduced the idea of concepts as soft detectors used to react to undetectable properties correlated with a system's detectable internal state. It may be interesting to extend the idea of resemblance in the manner that would make it applicable to cases where a detection mechanism is involved. Here the relevant facts about possible denotata would be undetectable properties and their connections with the system's intrinsic properties. The introduced notion of resemblance is general enough to deal with such cases.

Roughly speaking, the notion of vividness may be characterized in terms of a missing aspect that transforms iconic and causal representations of something for someone into the representations that constitute subject's acquaintance with the object.[21] The idea here is that the vehicle of representation must play a sufficiently distinguished role in the subject's thoughts that intend to be about some particular object *x*. Although this general description of vividness is sufficient for purpose of the paper it should be stressed that there is currently no agreement how to describe the notion of sufficiently distinguished role.[22] The important point, however, is this: each such theoretical description known to the author is consistent with the proposal introduced in the section 3.

Equipped with iconicity, causality and vividness we are now prepared to distinguish between various cognitive situations that involve agents. Let us use asterisks (Kaplan's corner quotes) as devices that indicate that we speak about complex representational structures. Now we have a contrast between:

(4) $\exists\alpha$ [Poirot BELIEVES *α is a thief* & $\boldsymbol{R}(\alpha$, the thief, Poirot)]

and

(5) $\exists\alpha$ [Poirot BELIEVES *α is a thief*] & $\neg\exists\beta$ [$\boldsymbol{R}(\beta$, the thief, Poirot) & Poirot BELIEVES *β is a thief*]

The first case corresponds to situations where Poirot's belief enables him to disclose who the culprit is. The second does not, on the other hand, as it fits very well to situations where Poirot thinks, for instance, that the thief is the person who broke the window but he has no idea who that person is.

[21] "If the name is such, that on the assumption there exists some individual *x* whom it both denotes and resembles, we should say that Ralph knows *x* or is acquainted with *x*, then the name is vivid" (Kaplan 1968, 201).

[22] One possible explication goes along the lines suggested by Robin Jeshion who proposes Significance Condition for singular thoughts (she uses here the terminology of mental files): "a mental file is initiated on an individual only if that individual is significant to the agent with respect to her plans, projects, affective states, motivations" (Jeshion 2010, 136). For a criticism of Jeshion account see (Geirsson 2017) where the author argues that a better alternative to Significance Condition is the condition of conscious attention.

Roughly speaking, the exportation is warranted if (4) is true of some particular person.[23]

I shall argue below that the representational apparatus introduced by Kaplan can deal with *de se*, *de hic* and *de nunc* beliefs provided that it is treated *experimentally*, that is as enabling various modifications of Kaplan's original idea, modifications like relaxing conditions for (certain kinds) of representations, putting new constraints on representations, augment its formal structure etc. Such an experimental approach, I think, fits well with Kaplan's declarations (cf. Kaplan 1968, 204).

3. Extending Kaplan's account

To start, I shall consider three modifications of Kaplan's representational approach.

The first one is that of *enriched relativization*. It seems that particular acts of representing involve a setting or a situation in which the representation takes place. This setting or situation, just like all the others, comprises a potentially infinite number of aspects or properties. Some of these aspects are relevant for the particular act of representing, while others are not. I would like to suggest that *de se*, *de hic* and *de nunc* beliefs are ones that engage particular *kinds* of such aspects: the agent that plays a particular role in the situation, its time and its location. In fact, the agent has already explicitly been featured as an argument of the representational relation—the idea, in a sense, is just to generalize this relativization into other aspects.

The second adjustment is that of *situation dependence*. Mere relativization (even enriched) does not mean much and, in particular, it does not entail that there exists any sort of dependence between the relevant aspect of the situation and the object of the representation. However, we may independently consider the possibility of such a dependence. In general, we may single out a class of situations where the object of the representation

[23] Due to space limitations, over the course of presentation I am ignoring several features of Kaplan's approach like the issue of the relative and gradable nature of vividness, the role of type-token distinction etc.

is a function of this or another aspect of the setting. If c is such a circumstance, setting or situation we may replace our original relation $\boldsymbol{R}(\alpha, x, y)$ with $\boldsymbol{R}(\alpha, x, y, c)$ (meaning: α is a representation of x for y at c) and consider the class of cases for which $\boldsymbol{R}(\alpha, f(c), y, c)$. In the cases we are interested in: $c = \langle a, t, l \rangle$, where a is the agent, t is the time and l is the place or location and the functional dependence is therefore of the simplest kind, as the object represented *is identical with* one of the aspects of situation (that is either with the agent a of the situation, the time t or the location l). Hence we may consider three general reflexive cases: the first in which we deal with a *representation of an agent* for *the agent* (represented as $\boldsymbol{R}(\alpha, a, a, \langle a, t, l \rangle)$), the second where we deal with a *representation of a time or instant at the time or instant* (represented as $\boldsymbol{R}(\alpha, t, a, \langle a, t, l \rangle)$) and the third in which we deal with a *representation of a location or place at the location or place* (represented as $\boldsymbol{R}(\alpha, l, a, \langle a, t, l \rangle)$).

The third idea is that of *uniqueness*. Some representations may be unique in the following sense: the change in the relevant aspect of the situation or setting requires a change of the vehicle of representation. Uniqueness basically means that a representation cannot occur if the circumstances are not of the appropriate kind or, alternatively, that it is unavailable for subjects who are not in the situation of the required kind. Due to that we may call such representations *elusive*. The concept of situations' relevant aspect differs for distinct kinds of indexical beliefs. Roughly speaking, in the case of *de se* attitudes it is the agent of the situation, in the case of *de nunc* attitudes it is the time of the situation, in the case of *de hic* attitudes it is the location of the situation.

The concept of *uniqueness* may sound familiar to those who endorse the idea of hybrid names (cf. Künne 1992; Textor 2007; Textor 2015). Hybrid names are names consisting of linguistic expressions and circumstances in which linguistic expressions occur. Although hybrid name theorists differ with respect to the question about the exact role of circumstances in reference fixing, they all agree that they are partially individuated by the circumstances in question. In the case of hybrid names, one is entitled to say that "[…] only utterances of 'I' *by the same person,* and only *simultaneous* utterances of 'the present moment' are occurrences of the same hybrid proper name" (Künne 1992, 725). Parallel considerations apply to unique

representations: *de se* representations may re-occur (that is they are occurrences of the same representation) only if the agent remains fixed, *de nunc* representations may re-occur only if the time remains fixed, *de hic* representations may re-occur only if the location remains fixed. However, there are also important differences between hybrid names and unique representations. These differences do not show that the two concepts cannot be combined within a single theory of representations or hybridity, they show rather that combining them would require substantial enrichment of one concept with the features of the other. Firstly, unique representations do not have to be linguistic or even partially linguistic.[24] They have to be mental representations, whatever the latter are. Hybrid proper names, on the other hand, essentially contain linguistic and circumstantial components. Secondly, unique representations of a particular kind (*de se*, *de nunc* or *de hic*) are representations that are iconic, causal and vivid. Nothing in the idea of hybrid name presupposes the three aspects. Thirdly, hybrid names are mereological sums of two kinds of components that both have properties determinative for the reference of a hybrid name. The first, the linguistic component, is endowed with meaning that tells how circumstances relate to the referent of the hybrid name, the second, the circumstantial component, saturates the meaning and gives reference. In contrast to hybrid names, unique representations do not have a special mereology: they may possibly be complex but have no referentially relevant parts or components. Fourthly, hybrid names as containing circumstances as parts are intrinsically functional upon situations.

Another theory that embraces the concept of uniqueness has been proposed by Peacocke (1981) and developed later by Higginbotham (Higginbotham 2003), see (Capone 2016, 232–36). Peacocke introduced the idea of special modes of presentations (self_x) and (now_t) that might be, respectively, *solely* constituents of the thoughts of a person x when x is thinking about himself or herself and constituents of the thoughts entertained at the time t that concern the time t. Higginbotham develops a view according to which sentences that have a *de se* interpretation employ (self_x) at the level

[24] On the other hand, they have to be semiotic in the full-blooded sense (cf. Konderak 2017).

of logical form.[25] The difference between the view defended in this paper and the views of Peacocke and Higginbotham is, firstly, that the former makes no use of the concept of mode of presentations and makes no use of the notion of sense (or its Carnapian counterpart—*intension*). As I shall explain below at the level of semantic interpretation belief reports state that there is a relation between an agent and a mental propositional representation, not between an agent and a proposition or a Fregean thought. Secondly, Peacocke-Higginbotham's idea of *de se* mode of presentation is not committed to the special epistemology of representation developed in this paper. The two differences, however, should not obscure the similarities between the approaches. One might, for instance, treat the account sketched in this paper as psychologized version of Fregean account as well as consider supplementing the Peacocke-Higginbotham's views with the special epistemology of representation. This will, actually, meet Peacocke's demand that one of the crucial tasks of the philosophy of mind is to say what the capability of thinking of objects under particular modes of presentations amounts to (Peacocke 1981, 194).

We are now in a position to define a minimal generic notion of indexical representation ("minimal" because it comprises just three aspects of the situation, "generic" because it uses the general concept of appropriateness that becomes definite for particular sorts of indexical beliefs (see below)):

For all α, x, y and $c = \langle a, t, l \rangle$ (for some agent a, time t and location l):

$\boldsymbol{RI}(\alpha, x, y, c)$ (= α indexically represents x for y in a situation c) iff

1. $\boldsymbol{R}(\alpha, x, y, c)$ (= α represents [in the sense of Kaplan] x for y in a situation c)
2. $x = f(c)$ (*situation dependence*)
3. for all c': if c is appropriately different from c', then it is necessary that: $\neg\boldsymbol{R}(\alpha, x, y, c')$ (*uniqueness*)

[25] Higginbotham stresses that there are differences between his and Peacocke but since they both embrace uniqueness of *de se* mode of presentation they are of secondary importance for this paper.

The concept of *being appropriately different* is intended to capture the fact that the indexical nature of representation does not mean a sensitivity to the change of the situation as a whole but rather a sensitivity to the change in one of its aspects (its agent, its time or its location). Two situations may be different but, if the difference is not the effect of the change in the appropriate aspect of the situation, the vehicle of representation might remain unchanged (an agent might clearly have constant *de se* representation of himself or herself at different times or at different locations).

I assume here a notational distinction between the subject of representation and the agent of the situation but one may, for the sake of notational simplicity, reduce $\boldsymbol{R}(\alpha, x, y, \langle a, t, l\rangle)$ to $\boldsymbol{R}(\alpha, x, \langle a, t, l\rangle)$ if she keeps in mind the assumption that the subject that represents and the agent of c are, in fact, identical. I will follow this convention below but the reader is kindly asked to bear in mind that the subject of the representation is always indirectly represented in $\boldsymbol{R}(\alpha, x, \langle a, t, l\rangle)$.

Corresponding notions of *de se*, *de hic* and *de nunc* representations may be defined as follows:

(Representing *de se*)
For all α, x, a, t, l:
$\boldsymbol{RS}(\alpha, x, \langle a, t, l\rangle)$ (= α is a *de se* representation of x at the situation involving the agent a, the time t and the location l iff

1. $\boldsymbol{R}(\alpha, x, \langle a, t, l\rangle)$ (= α represents x at the situation involving the agent a, the time t and the location l)
2. $x = a$
3. for all a': if a' is different from a, then—for each t^* and l^*—it is necessary that: $\neg\boldsymbol{R}(\alpha, x, \langle a', t^*, l^*\rangle)$

(Representing *de hic*)
For all α, x, a, t, l:
$\boldsymbol{RH}(\alpha, x, \langle a, t, l\rangle)$ (= α is a *de hic* representation of x at the situation involving the agent a, the time t and the location l) iff

1. $\boldsymbol{R}(\alpha, x, \langle a, t, l\rangle)$ (= α represents x at the situation involving the agent a, the time t and the location l)
2. $x = l$
3. for all l': if l' is different from l, then—for each a^* and t^*—it is necessary that: $\neg\boldsymbol{R}(\alpha, x, \langle a^*, t^*, l'\rangle)$

(Representing *de nunc*)
For all α, x, a, t, l:
$\boldsymbol{RN}(\alpha, x, \langle a, t, l\rangle)$ (= α is a *de nunc* representation of x at the situation involving the agent a, the time t and the location l) iff

1. $\boldsymbol{R}(\alpha, x, \langle a, t, l\rangle)$ (= α represents x at the situation involving the agent a, the time t and the location l)
2. $x = t$
3. for all t': if t' is different from t, then—for each a^* and l^*—it is necessary that: $\neg\boldsymbol{R}(\alpha, x, \langle a^*, t', l^*\rangle)$[26]

RI, ***RS***, ***RH*** and ***RN***, recall, are all defined in terms of the original Kaplanian relation of representation. Hence, they all meet iconicity, causality and vividness demands.

Now, just like *de re* belief reports can be analyzed along the lines depicted in (4) the reports similar to (2) (*A believes that she herself is F*) receive the following representation:

(6) $\exists\alpha$ [A BELIEVES *α is F* & $\boldsymbol{RS}(\alpha, \text{A}, \langle \text{A}, \mathbf{t}, \mathbf{l}\rangle)$]

and the reports like (3) (*A believes that she (but not that she herself) is F*) the following:

(7) $\exists\alpha$ [A BELIEVES *α is F* & $\neg\boldsymbol{RS}(\alpha, \text{A}, \langle \text{A}, \mathbf{t}, \mathbf{l}\rangle)$[27]

Due to the risk of boring my readers I will not go through the analogous cases of *de hic* and *de nunc* beliefs. Note that (7) can be made true by the entire class of different states of affairs that involve the psychological state of the agent A. Firstly, (7) is true if any of the three constitutive conditions for ***RS*** fails to obtain. This might happen, for instance, if α is not vivid, it may happen if it is not iconic or if it does not meet the causality demand. This embraces five possible kinds of cases if we assume that vividness requires iconicity and causality and even more if we deny that assumption.

[26] Points 1 and 2 of each definition may just be, of course, reduced to overtly reflexive $\boldsymbol{R}(\alpha, a, \langle a, t, l\rangle)$, $\boldsymbol{R}(\alpha, t, \langle a, t, l\rangle)$ and $\boldsymbol{R}(\alpha, l, \langle a, t, l\rangle)$ respectively.

[27] In cases where the agent has no disposition at all to regard himself or herself as being F (she suffers from crime related amnesia, for instance) the reading may be even strengthen to: $\exists\alpha$ [A BELIEVES *α is F* & $\forall\beta$ (A BELIEVES *β is F* $\Rightarrow$ $\neg\boldsymbol{RS}(\beta, \text{A}, \langle \text{A}, \mathbf{t}, \mathbf{l}\rangle)$] which entails (7).

Combining this with other cases where situation dependence or uniqueness fails we arrive at twenty-three *kinds* of situations that can make (7) true. This illustrates the fact that there is no simple way of connecting belief reports (at least some belief reports) and kinds of belief states that fit the belief reports in question. However, at the very same time the semantics of belief reports indicates features that every belief state must have in order to qualify as a potential truthmaker of the belief report.

The following bigger picture emerges from the analysis just presented. Beliefs, roughly, may be described as relations between agents and propositional representational structures. If we treat as criterial (for being a *de dicto* propositional attitude) the fact that the relation takes a propositional structure as an argument, then all the kinds of beliefs described above count as *de dicto*. *De re* beliefs, from that perspective, are the ones that involve relation to the propositional structure that contains at least one iconic, causal and vivid representation. Indexical beliefs are, consequently, the ones that involve the relation to the propositional structure containing iconic, causal and vivid representation that is, additionally, relativized, situation dependent and unique. In the case of *de se* beliefs this commits us to the well-known Fregean self-presentation demand as our definition of *de se* entails that "everyone is presented to himself in a special and primitive way, in which he is presented to no one else" (Frege 1956).

The theory sketched above differs in several aspects from the most popular approaches to indexical beliefs. Firstly, the approach is not committed in any way to special kinds of *de se*, *de hic* or *de nunc* or relativized contents (it is, however, committed to *representations of limited accessibility*, more about that below). In particular, it does not describe beliefs in the Lewisian manner, that is as relations between agents and properties. It also does not make use of centered worlds (cf. Stalnaker 2008) as it postulates a more fine-gained notion of belief described in terms of representations. Secondly, it differs also from the approach of John Perry as it does not introduce the special concept of a belief state that determines the proposition relatively to a situation. Here, however, the situation looks slightly different as the discussed notion of representation is clearly situation relative. If one thinks, therefore, that using a particular representation shapes the belief state of an agent, then the present account may be seen as a representationalist variant of Perry's approach.

The account presented offers a solution to all three problems of locating beliefs described by Feit and Capone. Firstly, *de se* beliefs differ from non-*de-se* beliefs in the kind of representation involved. In particular, akin *de re* representations are neither situation dependent nor unique. Secondly, the semantic problem is solved in two steps. One is by noting that there is no need to postulate a special semantic value of "I" or "she herself" other than the referent. Another is by noting that at the level of semantic representation the belief reports that contain such phrases explicitly make use of a special relation of representation. Thirdly, if we interpret the Geach problem as a question regarding the interpretation of our propensity to treat as cases of samebelieving situations, for example like the one where both Donald and Hilary think that they will win the next presidential elections,[28] then the corresponding belief reports are:

(8) $\exists\alpha$ [Hilary BELIEVES *α will win* & $\boldsymbol{RS}(\alpha, \text{Hilary}, \langle \text{Hilary}, \mathbf{t}, \mathbf{l}\rangle)$]

(9) $\exists\alpha$ [Donald BELIEVES *α will win* & $\boldsymbol{RS}(\alpha, \text{Donald}, \langle \text{Donald}, \mathbf{t}, \mathbf{l}\rangle)$]

It is, I think, clearly visible that both Hilary and Donald are related to representations that have not only isomorphic structure but also involve a single *kind* of representation. This, I think, suffices for explaining the samebeliving intuition in question. On the other hand, due to the uniqueness demand, they *cannot be related to the same representation of the relevant kind*. This naturally explains the opposite intuition.

4. Loosening the requirements put on indexical beliefs: five doubts

I conclude that the theory sketched above offers a promising analysis of some aspects of the problem of indexical belief. My plan now is to discuss some potential challenges that the theory must face.

[28] We must, of course, keep in mind that we also have an opposite propensity to treat such cases as the ones of non-samebelieving.

4.1. De re beliefs about oneself

John Perry, after discussing the proposals of Kaplan and Hintikka construing *de re* attitudes as special cases of *de dicto* attitudes, defied the idea of the applicability of such an approach to indexical beliefs on the following grounds:

> However well these proposals deal with other phenomena connected with *de re* belief, they cannot help with the problem of the essential indexical. They tighten the requirements laid down by the original proposal, but those were apparently already too restrictive. If in order to believe that I am making a mess I need not have any conceptual ingredient α that fits me, *a fortiori* I am not required to have one that is a vivid name of myself for me, or one that picks out the same individual in every possible world compatible with what I believe. (Perry 1979, 11)

Immediately Perry also provided the examples of *de re* beliefs of an agent about oneself that are not indexical (Perry 1979, 12). One goes as follows. Imagine that I see the reflection of the messy shopper in the mirror and I express my belief by saying "He is making a mess." By doing this I express a *de re* belief about myself. However, if I am not aware that this is my reflection I see, this *de re* belief is not a *de se* one.[29]

I have no doubts that such examples conclusively prove that the class of *de se* beliefs and the class of *de re* beliefs about oneself are distinct. Does this fact pose an insuperable challenge to the theory sketched above?

There can be little doubt that in such cases the demands of iconicity, causality, vividness and enriched relativization are met. What about situation dependence and uniqueness? If we equate the situation, as we did, with the triumvirate that consists of an agent, a location and an instant, then the situation dependence means that the representation involved in *de se* cases is a function of the agent of the situation. And this is clearly the case: accidentally or not, the object represented is identical with the agent that has the relevant belief. However, the uniqueness demand is not met by the

[29] Similar cases have been considered in the early 70's by Ernest Sosa (Sosa 1970, 893). Perry refers to such cases as cases of "belief of the person one happens to be" (Perry 1998), Recanati (Recanati 2007) calls such beliefs "accidental *de se*."

agents involved in the cases of accidental *de se* beliefs. *Other person* may also see the reflection of the messy shopper and utter "He is making a mess." I see no reason to claim that there is a difference in the kind of vehicle of representation between myself (having accidental *de se* belief) and such persons (having a regular *de re* belief about myself). Therefore, the presented account of *de se* beliefs is not undermined by the discussed examples.

Inspired by scenarios involving accidental *de se* beliefs Boër and Lycan (Boër and Lycan 1980) suggested that there is no real contrast between the case of indexical beliefs and the case of non-indexical *de re* beliefs. The reason for this is that one might consider scenarios involving non-indexical *de re* beliefs that strictly follow the structure of cases involving accidental *de se* attitudes. One can, for instance (cf. Boër and Lycan 1980, 448–50), consider the story of John who sees two distant persons. One is followed by a predator while the other looks perfectly safe. John dubs the first person 'Van' and the other 'Wilfrid.' In fact, Van is just a mirror reflection of Wilfrid and the mirror reflection reveals the predator that cannot be seen in the flesh. It seems that we have here two *de re* beliefs about single person that correspond to accidental and regular *de se* beliefs.

The framework sketched above, I believe, enables to see the difference between the two cases. If Van-Wilfrid scenario involves unique representations, then *both* cases can only be perceptually unique (Van is the men that looks thusly from here now while Wilfrid is the men that looks differently from here now). Therefore, there is a symmetry between perceiving Van and perceiving Wilfrid as the change in the relevant aspect of the perceptual situation enforces the change in the representation itself (the change of the Van-like representation into the Wilfrid-like representation, for instance). However, as it has been noted above, *no parallel symmetry* in the cases of regular and accidental *de se* exists as accidental *de se* cases do not meet the uniqueness requirement with respect the relevant aspect of the situation.

4.2. Selfless knowledge

A different challenge is, I think, posed by the eventuality that "in order to believe that I am making a mess I need not have any conceptual ingredient α that fits me." The idea fits well with the cluster of theories developed in other papers of John Perry (cf. Perry 1986; and Kim 2010, 60–62)

where the ideas of thought without representation and the essentially selfless knowledge are introduced and discussed. Firstly, Perry argues there that it is, in principle, possible to have thoughts about something without having a representation of that thing. Secondly, he notes that many thoughts that play a motivational role in causing our actions seem to be selfless. Now, if we enable thoughts without representation and the idea of intrinsically selfless knowledge, then the possibility that at least *some de se* thoughts involve no representation of the subject may look compelling.[30] And if it is compelling—one may suggest—an adequate account of *de se* beliefs cannot appeal to representations regardless of their elusiveness.

We have, therefore, two competing pictures of locating beliefs: one committed to elusive representations and another committed to non-self-representationalism. I think that the phenomenon of essential indexicality together with other broadly accepted observations regarding the explanation of human actions does, in general, favor the representationalism defended in this paper. Suppose that I am just about to write down the word "self." In order to do that I have to press a certain sequence of keys on the keyboard *in front of me*. The keys are thusly located *with respect to me*, the sequence of little key-pressing actions I have to perform is determined partially by *the goal I have*. All of these shape the action I actually perform. The objects involved (the keyboard, keys and the sequence of letters qua the intended result) might be conceived as entities that play *agent-relative roles* (cf. Perry 2000, 326). I clearly may have knowledge of all such things without representing them as accordingly related to me. However, can such knowledge be solely responsible for the explanation of action in cases where indexical beliefs seem necessary? Consider our absent-minded referee again. It is an essential feature of Monica's belief that by gaining the knowledge about the authorship of the paper she is gaining the knowledge that is *explicitly about her*. This is why we are in the position to dismiss all corresponding non-egocentric explanations that are not accompanied by the additional belief that she herself plays a corresponding non-egocentric role. This is what makes her belief essentially indexical. Hence, there are reasons

[30] Naturally, some *de se* thoughts may engage the representation of the subject but the very fact that the opposite is possible entails that containing a representation (of any sort) is not obligatory for being a *de se* thought.

to think that the classes of selfless thoughts and indexical thoughts are disjoint. Due to this fact one cannot argue against elusive representations on the basis of the possibility of thoughts without self-representation.

4.3. En report requirement

I should like now to consider a different kind of worry. The approach presented above is based on David Kaplan's ideas and in his later writings Kaplan himself expressed skepticism regarding this stage of his philosophical development (cf. Kaplan 1985, 326). He also suggested that his content/character distinction constitutes a more promising framework for dealing with motivational role of beliefs. Kaplan's skepticism concerns primarily his conception of *de re* propositional attitudes and, as such, questions the bigger picture of beliefs described at the end of the section three. A natural reaction to Kaplan's skepticism is noting that even though the bigger picture may be incorrect, this does not prove that the *en report* requirement (which consists in the demands of iconicity, causality and vividness) is not necessary for indexical beliefs. In fact, it is difficult to imagine the cases in which one has a locating belief without being *en report* with its object. So, even if Kaplan is right about some *de re* cases, his observation does not extend to *de se*, *de hic* and *de nunc* ones.

An alternative reply may stress a difference between *expressing a singular proposition* and *believing* (and eventually *grasping*) *it*. It is, I think, by no means accidental that Kaplan's example concerns assertion (not belief). Asserting something undoubtedly involves a public language with all its reference-fixing aspects and ubiquitous linguistic conformism. Gareth Evans noted once that expressions are often used *deferentially*, that is with the intention of conforming to the use of others (Evans 1973, 205). In such cases a person may clearly express a proposition even though her beliefs may gave her an inadequate idea about the object she is referring to and, consequently, about the proposition she actually expresses. This, among other things, enables her to express a proposition without entertaining it. The first child example used by Kaplan (*Dthat ('the first child to be born in the twenty-first century') will be bald*) might be interpreted as an interesting case of a similar type: we expressed a singular proposition, which we were not yet capable of grasping (in 1978 when the example was given).

The difference between Evans deferential uses and the Kaplan example would be that in the latter case we conform to the use of a description in the whole linguistic community while instrumentally employing this usage as a demonstrating procedure. Nothing similar applies to deferential uses. Despite this difference, however, both cases may potentially lead to the assertion of the proposition that *p* without belief that *p*.

I do not want to favor one of the two possible replies just presented. The important point is this: Kaplan's worry regarding his own theory does not automatically extend to self-locating beliefs, besides any such extension is highly problematic.

4.4. Propositions of limited accessibility

We have seen that one of the most important features of the described approach to self-locating beliefs is the fact that it postulates a class of unique representations—ones that are not repeatable if the change in the situation affects it in a particular manner. This may suggest that our theory is committed to *propositions of limited accessibility* (PLA).

The idea of propositions of limited accessibility was introduced by John Perry who described it as claiming that:

> […] there is a class of propositions which can only be expressed in special circumstances. In particular, only I could express the proposition I expressed when I said 'I am making a mess.' Others can see, perhaps by analogy with their own case, that there is a proposition that I express, but it is in a sense inaccessible to them. (Perry 1979, 16)

Although Perry described theories that are committed to PLA as offering a consistent solution to the problem of essential indexicality, he expressed his skepticism towards them. Firstly, he pointed out that one needs additional metaphysical reasons to enable PLA. Secondly, he suggested that such reasons must somehow presuppose the ontology of „private perspectives" (Perry 1979, 16). This, in turn, questions our belief in „the common actual world" (Perry 1979, 16).

It is, therefore, important to stress that the theory sketched in this paper is literally committed not to *propositions* of limited accessibility but rather

to *representations* of limited accessibility (RLA). Now, we clearly do not have to equate PLA and RLA. Here are examples of three theories that are may embrace the distinction. Firstly, the representational theory of the mind, for instance, offers a theoretical framework that describes attitudes as directly related to mental representations and indirectly (*via* mental representations) to contents (including the propositional ones). Any account that sympathizes with the representational theory of the mind might, therefore, postulate RLA but deny any sort of commitment to PLA. Secondly, similar remarks apply to theories that approach the concept of proposition in a classificatory way.[31] Thirdly, despite all the differences between hybrid names and unique representations there is one important parallel between the two ideas. Both unique representation and hybrid names are not contents but rather means of expression of the content. By distinguishing between propositions and representations the theory sketched in this paper assumes that objects of attitudes (representations) and semantic values of sentences (propositions) are different entities.

Since Stoic times there have been two main traditions of interpreting the concept of proposition. The first, represented for instance by Bolzano and Frege, construed them as abstract, mind independent entities. The second, represented for instance by medieval nominalists, construed them as mind dependent entities, often expressions of the mental language. Neither of the two theories, I think, would be happy to accept propositions of limited accessibility and neither would have a problem with representations of limited accessibility. The only theory that may be somehow committed to PLA is the one that equates propositions with representations and, at the very same time, treats them as abstract or, at least, completely mind independent entities (like states of affairs or situations). Such a theory would endow cognitive subjects with the capacity of having thoughts or attitudes that are *direct* relations to such mind independent entities. If we interpret

[31] Here John Perry stresses the importance of the representation-proposition distinction: "My conception […] is that propositions are abstract objects we use to classify cognitive states, especially states that involve representations. Propositions are not representations, but abstract objects that we use to classify states and events by the requirements of their truth (or some other form of success imposed on the rest of the world)" (Perry 2012, 27).

the theory presented in this paper along such lines, then it is unquestionably committed to representations qua propositions of limited accessibility, However, I see no rationale for interpreting or developing it in that manner.

4.5. The account is "too fine-grained"

One might worry that the view defended in this paper provides too fine-grained criteria of belief individuation. The uniqueness constraint prohibits that indexical beliefs can be entertained in situations that appropriately differ from the ones where they are actually entertained. However, it seems that there is a sense in which my thought that I have two hands now (entertained at the time t) and my thought that I have two hands now (entertained at the time t′ different from t) have the same content.[32] One aspect of the problem has already been addressed above: the case of two now-thoughts seems to be a temporal counterpart of the example used to illustrate Geach problem. The analysis of the cases, therefore, is similar: on the one hand, we may explain the intuition of sameness of beliefs by the fact that the two beliefs are isomorphic. There is, however, another aspect of the problem that deserves attention. If we follow strictly the observation that our theory treats beliefs as relations to mental representations, then the following *possibility* emerges: distinct mental representations might be linked to a single propositional content. If we embrace this possibility, there are at least two levels at which beliefs might be individuated. One is the level of propositional representations. At this level, two *de nunc* thoughts

[32] The anonymous referee asked me to consider examples of two first person thoughts that I have hair today entertained on two consecutive days. Since my account says nothing about *hoc die* attitudes I decided to change the example to two *de nunc* thoughts. However, the point raised by the referee motivates another important question: should the account sketched in this paper be generalized to other sorts of indexical thoughts (today-thoughts, yesterday-thoughts, there-thoughts, you-thoughts etc.). My initial reaction is that is should not as the thoughts in question seem not to require uniqueness (I agree here with Frege's famous observations concerning "today" and "yesterday," I denied this generalization for demonstrative also when discussing the case of accidental *de se*). However, I do not have at the present moment the general criterion of demarcation and I would like to stress that this important issue requires further studies.

entertained at different times might be similar but never strictly type-identical. Another is the level of propositional content. Here, despite being related to two distinct propositional representations, two *de nunc* thoughts entertained at different times might have the same propositional content. We have to be very careful here, however, when it comes to the analysis of particular examples. On the one hand, we might follow Prior's intuitions (Prior 1970) that redundancy theory of the present (or redundancy theory of *being here*) correctly captures the content of now-thoughts and here-thoughts so my thought that I have two hands now is nothing more than the thought that I am having two hands. On the other hand, a similar approach to *de se* thoughts seems superficial and controversial.[33] That having been said, at the present moment I prefer to treat this application of the propositional content/proportional representation distinction as the signpost of one of several ways in which the theory described in this paper might be developed.

5. Conclusions

I have argued that a modification of Kaplan's approach to *de re* belief reports can be consistently extended to cover various cases of indexical beliefs. The resulting theory postulates a special class of unique circumstance-dependent representations. However, it is neither committed to propositions of limited accessibility, nor to any sort of special indexical contents. The theory has numerous merits. Firstly, it offers a solution to psychological and semantic problems of indexical beliefs. Secondly, it enables us to solve the Geach problem. Thirdly, it has the resources to distinguish cases of regular *de se* and accidental *de se* beliefs as well as to address Boër and Lycan worries that there is no important difference between indexical and non-indexical *de re* attitudes. Fourthly, it gives a general and consistent picture of *de dicto, de re* and indexical attitudes. In particular, it indicates the similarities and dissimilarities between the three classes of attitudes. Last but not least, it seems to promise an area for further interesting

[33] Nevertheless, it still might be developed in the framework of egocentric logic developed by Prior (cf. Prior 1968, 135–44).

philosophical developments. The account seems, for instance, to fit very well to the representational theory of the mind, the theories that appeal to the idea of representations having forms (like the mental orthography approaches), it may also be perceived as supporting the view of concepts as mental representations (cf. Margolis and Laurence 2007). All these merits, I think, support the theory as being worthy of further investigation and development.

Acknowledgements

Ideas presented in this paper were discussed at Semantic Content and Conversational Dynamics workshop (Barcelona), Philang 2017 (Lodz), South American Congress of Analytic Philosophy (Campinas), European Congress of Analytic Philosophy (Munich), during my short visit at the University of Konstanz and at the meeting of Institute of Philosophy of the Czech Academy of Sciences. I would like to thank the audiences and the anonymous reviewers of *Organon F* for insightful comments and suggestions.

Funding

The work on this paper was funded by National Science Center, Poland, grant under award number 2013/11/B/HS1/03947.

References

Boër, Steven E., and Lycan, William G. 1980. "Who, me?" *The Philosophical Review* LXXXIX (3): 427–66. https://doi.org/10.2307/2184397

Capone, Alessandro. 2016. *The Pragmatics of Indirect Reports. Socio-Philosophical Considerations*. Dordrecht: Springer. https://doi.org/10.1007/978-3-319-41078-4

Cappelen, Herman, and Dever, Josh. 2013. *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199686742.001.0001

Castañeda, Hector N. 1966. "'He': A Study in the Logic of Self Consciousness." *Ratio* 7: 130–57.

Chierchia, Gennaro. 1989. "Anaphora and Attitudes De Se." In *Semantics and Contextual Expressions*, edited by Renate Bartsch, Johan van Benthem, and Peter van Emde Boas, 1–31. Dordrecht: Foris.

Ciecierski, Tadeusz. 2016. "Linguistic Criteria of Intentionality." *Studies in Logic, Grammar and Rhetoric* 46 (1): 35–58. https://doi.org/10.1515/slgr-2016-0032

Corazza, Eros. 2004. *Reflecting the Mind.* Oxford: Oxford University Press. https://doi.org/10.1093/019927018X.001.0001

Cresswell, Max. 1985. *Structured Meanings.* Cambridge (MA): The MIT Press.

Evans, Gareth. 1973. "The Causal Theory of Names." *Proceedings of the Aristotelian Society, Supplementary Volume*s 47: 187–225. https://doi.org/10.1093/aristoteliansupp/47.1.187

Feit, Neil, and Capone, Alessandro. 2013. "The Problem of De Se Attitudes." In *Attitudes De Se. Linguistics, Epistemology, Metaphysics,* edited by Neil Feit and Alessandro Capone, 1–25. Stanford: CSLI Publications.

Frege, Gottlob. 1956. "The Thought: A Logical Inquiry." *Mind* 65: 289–311.

Geirsson, Heimir. 2017. "Singular Thought, Cognitivism, and Conscious Attention." *Erkenntnis* 83 (3): 1–14. https://doi.org/10.1007/s10670-017-9905-3

Gillon, Brendan S. 1990. "Ambiguity, Generality, and Indeterminacy: Tests and Definitions." *Synthese* 85: 391–416. https://doi.org/10.1007/BF00484835

Grabarczyk, Paweł. 2016. Concepts as Soft Detectors. *New Ideas in Psychology* 40: 86–93. https://doi.org/10.1016/j.newideapsych.2015.03.001

Higginbotham, James. 2003. "Remembering, Imagining, and the First Person." In *Epistemology of Language*, edited by Alex Barber, 496–535. Oxford: Clarendon. https://doi.org/10.1093/acprof:oso/9780199239313.003.0012

Jaszczolt, Kasia. 2013. "Contextualism and Minimalism on De Se Belief Ascription." In *Attitudes De Se. Linguistics, Epistemology, Metaphysics,* edited by Neil Feit and Alessandro Capone, 69–103. Stanford: CSLI Publications.

Jeshion, Robin. 2010. "Singular Thought: Acquaintance, Semantic Instrumentalism and Cognitivism." In *New Essays on Singular Thought*, edited by Robin Jeshion, 105–40. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199567881.003.0005

Kaplan, David. 1968. "Quantifying in." *Synthese* 19 (1-2): 178–214. https://doi.org/10.1007/BF00568057

Kaplan, David. 1985. "Dthat." In *The Philosophy of Language*, edited by Aloysius Martinich, 315–28. Oxford: Oxford University Press.

Kim, Hanseung. 2010. "Two Notions of Indexicality." *Semiotica* 180 (1): 47–67. https://doi.org/10.1515/semi.2010.030

Konderak, Piotr. 2017. "The Conscious Semiotic Mind." *Studia Semiotyczne* 31 (1): 67–89. https://doi.org/10.26333/sts.xxxi1.05

Kripke, Saul. 2011. "Unrestricted Exportation and Some Morals for the Philosophy of Language." In *Philosophical Troubles,* 322–50. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199730155.001.0001

Künne, Wolfgang. 1992. "Hybrid Proper Names." *Mind* 101 (404): 721–30. https://doi.org/10.1093/mind/101.404.721

Lewis, David. 1979. "Attitudes *De Dicto* and *De Se*." *Philosophical Review* 88 (4): 513–43.

Margolis, Eric, and Laurence, Stephen. 2007. "The Ontology of Concepts—Abstract Objects or Mental Representations?" *Noûs* 41 (4): 561–93. https://doi.org/10.1111/j.1468-0068.2007.00663.x

Peacocke, Christopher. 1981. "Demonstrative Thought and Psychological Explanation." *Synthese* 49 (2): 187–217. https://doi.org/10.1007/BF01064298

Perry, John. 1979. "The Problem of the Essential Indexical." *Noûs* 13 (1): 3–21.

Perry, John. 1986. "Thought without Representation." *Proceedings of the Aristotelian Society*, Supplementary Volumes 60: 137–66. https://doi.org/10.1093/aristoteliansupp/60.1.137

Perry, John. 1998. "Myself and I." In *Philosophie in Synthetischer Absicht*, edited by Marcelo Stamm, 83–103, Klett-Cotta. Reprinted in (Perry 2000, 325–39).

Perry, John. 2000. *The Problem of Essential Indexical and Other Essays*. Stanford: CSLI Publications.

Perry, John. 2013. "Self-Locating Beliefs." In *Attitudes De Se. Linguistics, Epistemology, Metaphysics,* edited by Neil Feit and Alessandro Capone, 385–408. Stanford: CSLI Publications.

Peterson, Philip L. 1997. *Fact, Proposition, Event*. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-015-8959-8

Prior, Arthur N. 1968. *Papers on Time and Tense*. Oxford: Oxford University Press.

Prior, Arthur N. 1970. "The Notion of the Present." *Studium Generale* 23: 245–48.

Recanati, François. 2007. "Imagining *De Se*." https://halshs.archives-ouvertes.fr/file/index/docid/160757/filename/de_se_imagining11.pdf

Rudnicki, Jakub. 2019. "Centred Propositions, what is Asserted, and Communication." *Theoria*. https://doi.org/10.1111/theo.12215

Schlenker, Philippe. 2011. "Indexicality and De Se Reports." In *Semantics. An International Handbook of Natural Language Meaning.* edited by Klaus von Heusinger, Claudia Maienborn, and Paul Portner, 1561–1604. Berlin: Mouton de Gruyter.

Sells, Peter. 1987. "Aspects of Logophoricity." *Linguistic Inquiry* 18: 445–79.

Stalnaker, Robert C. 1981. "Indexical Belief." *Synthese* 49: 129–51. Reprinted in: (Stalnaker 1999, 130–49) https://doi.org/10.1007/BF01063917

Stalnaker, Robert C. 1999. *Context and Content*. Oxford: Oxford University Press. https://doi.org/10.1093/0198237073.001.0001

Stalnaker, Robert C. 2008. *Our Knowledge of the Internal World*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199545995.001.0001

Sosa, Ernest. 1970. "Propositional Attitudes *De Dicto* and *De Re*." *The Journal of Philosophy* 67 (21): 883–96. https://doi.org/10.2307/2024040

Textor Mark. 2007. "Frege's Theory of Hybrid Proper Names Developed and Defended." *Mind* 116 (464): 947–81. https://doi.org/10.1093/mind/fzm947

Textor, Mark. 2015. "Frege's Theory of Hybrid Proper Names Extended." *Mind*, 124 (495): 823–47. https://doi.org/10.1093/mind/fzv032

Weber, Clas. 2013. "Centered Communication." *Philosophical Studies* 166 (1): 205–23. https://doi.org/10.1007/s11098-012-0066-6

Vendler, Zeno. 1967. *Linguistics in Philosophy*. Ithaca: Cornell University Press.

Verdejo, Victor. 2017. "Explaining Public Action." *Topoi*. https://doi.org/10.1007/s11245-017-9524-2

Organon F 27 (3) 2020: 377–394
https://doi.org/10.31577/orgf.2020.27305

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

How Is Vicarious Feeling Possible? In Defense of Reactive Attitudes

Sunny Yang*

Received: 16 June 2019 / Accepted: 12 February 2020

Abstract: My aim in this paper is to illuminate the question of how vicarious feeling is possible, by advancing our understanding of vicarious emotions. I address this problem by classifying the reactive attitude into two categories: the vicarious, and the self-reactive. I argue that guilt is constitutively tied to personal responsibility and that the appropriateness of vicarious feeling of group harm derives from a reflection on the appropriateness of our own reactive attitude, that is, vicarious reactive attitude, e.g., indignation or outrage.

Keywords: Collective guilt; indignation; reactive attitude; responsibility; vicarious feeling.

1. Vicarious feeling

Vicarious feeling is a feeling that is experienced on behalf of someone else's action. In order to explore this emotion, let us first of all, compare some emotions, especially, pride and guilt. Both are obviously related to the self. Vicarious pride can be said to be the propensity to imagine oneself in the position of a loved one and thereby to feel, from that perspective, the pleasure that the qualities possessed by the loved one command. It seems

* Hankuk University of Foreign Studies
Minerva Liberal Art College, Hankuk University of Foreign Studies, 81, Oedaero, Mohyeon, Choingu, Yongin, Gyeonggi, 17035, South Korea
sunnyyang@hufs.ac.kr

possible to experience a feeling in relation to one's loved ones, as when we feel after seeing them as extensions of ourselves. For example, when my daughter achieves something, it is easy for me to feel proud given that she is a part of me, and thereby to feel proud on her behalf. The way that admiration or love can lead to pride is by assimilation or identification. It is difficult to imagine how I could come to attribute pride to you without recognizing in you pleasure at something you are related to. Hence, it seems that vicarious pride in a loved one's achievements requires genuine self-consciousness. We can call this view the cognitivist theory of emotion. It involves propositional attitudes which are complex and contentful.[1] If this kind of recognition is necessary, then what I am experiencing cannot be a primitive, pre-cognitive form of empathy or of the sentiment of pride. Then one might argue against such a purely cognitivist stance—they might argue that self-consciousness and recognition in others does not need simulation theories or content, but just some reflections of X. This type of empathy or sympathy in other circumstances seems to necessarily involve a kind of cognition.

In this respect, cognitivists say that when I am proud of my beautiful house it is because of my *belief* that the object is mine. In being proud of my beautiful house, I first of all must *believe* that it is valuable; secondly, in order for the feeling to play a role I must *believe* the house to be in some way connected with me. G. Taylor calls those two beliefs 'explanatory' and 'indentificatory,' respectively (Taylor 1985, 27). The 'explanatory' belief just explains the relation between the valuable things and the person, whereas 'indentificatory' belief refers to something 'closely' related to the person who feels pride. Thus, according to Taylor, "a person may hold the requisite explanatory beliefs and yet not feel proud." "She may regard her beautiful house as a most desirable possession but may not regard this as reflecting on her own worth" (Taylor 1985, 34). Thus, in order to feel pride, there must be indentificatory belief that "the agent regards the desirable

[1] The question whether the theory of emotion should count emotion as cognitive or not is the question whether cognitive elements, such as belief and judgment, are conceptually necessary or not for having emotion. With regard to 'cognition,' there has been a lively debate within the 'cognitivist' camp whether the type of cognition in question is better thought of as belief, thought, judgment, or something else. For a more detailed discussion on the cognitivist theory of emotion, see (Yang 2016).

thing as something she herself has brought about." That is, she must regard the information given by explanatory beliefs as contributing to her worth. This is, according to Taylor, a sufficient condition for pride. But if we accept this view we cannot explain the following case: in the case of the triumph of the team which I support, pride may involve 'explanatory belief,' but not involve 'indentificatory belief,' since I cannot regard the team's victory as one that I myself brought about. Thus, in this respect, some people say that the pride in the triumph of the team does not derive from belief but from my *thinking of* the team's victory *as* mine (as argued in Yang 2016). Yet one might wonder how this is different from merely imagining the victory as his. He can, in some sense, imagine the Spanish football victory as his, but it does not make him proud of it.

Another difficulty faced by cognitivism in explaining vicarious feelings is related to the phenomenological features of such emotions. It seems to be possible to feel guilty, for example, by seeing someone as an extension of oneself. A mother can feel guilt for a wrong committed by her son. If it is a necessary feature of any emotional state of guilt that it has certain phenomenological features, such as feelings of discomfort and distress, then the question arises, whether someone can truly have a vicarious feeling when they do not affectively respond. In light of this, cognitivists might question the assumption that phenomenology is constitutive of the emotions. For example, Margaret Gilbert says:

> When I say to you 'I feel great remorse' must I be saying something false unless there are pangs or the like in the background? On the face of it, I need not be saying something false. Note that some apparently equivalent expressions do not use the term 'feel' at all: 'I am full of remorse'; 'I am truly remorseful' (Gilbert 2000, 135).

If this were right, we can say that a shared feeling of guilt is possible on the basis of cognitivism on emotion, for example, a judgmentalist view, according to which an emotion's essential element is a judgment, while phenomenal feeling just accompanies it. There are variations on cognitive theory of emotion. In this paper I focus on Robert Solomon's (1993) and Martha Nussbaum's (2001) views, for the sake of criticizing Gilbert's position on collective emotion. This is because Gilbert seems to endorse Nussbaum's view in analysing collective guilt, although she does not specifically focus

on emotion theory itself. According to Solomon (1993) and Nussbaum (2001), an emotion can be analysed solely in terms of beliefs, or evaluative judgments. Furthermore, they argue that non-cognitive elements or bodily feelings are not necessary or sufficient elements of an emotion. Rather they take evaluative judgments to be the constituent parts of an emotion. Nussbaum seems to try to explain how emotions can be violent, stormy, fading, suffocating etc. in purely cognitive terms without reference to (bodily) feelings. Given this, the main objection is that it overlooks the affective aspect of emotion. It fails to do justice to a person's emotional state such as their being upset, perturbed, agitated or moved. The judgmentalists also acknowledge that most emotions are accompanied by physiological changes and their feelings. Moreover, they think that these physiological changes are caused by the combination of the appropriate belief and desire. Yet, they believe that physiological changes are just a *symptom* of a given emotion, but are not necessary for the concept of emotion.

Following cognitivist view on emotion, Gilbert argues that a shared feeling of guilt is a function of the joint commitment to form a unified subject that expresses guilt. But how could there be a unified subject that feels guilt? In order to understand this idea, let us look closely at Gilbert's cognitive view of collective guilt.

2. Gilbert's cognitivist view of collective guilt

Gilbert examines shared emotions through her pioneering view on collective emotions, and she presents an account of guilt in groups, specifically, in her article "Collective guilt and collective guilt feelings" (2002).

Now, one might say that only individuals have feelings, whereas groups do not have consciousness and therefore cannot feel anything either. Gilbert ignores this problem by adopting a strong cognitivist line on emotions. She argues that these emotions directed at collectives cannot be analyzed in terms of individual guilt. Thus, she argues that in order to feel guilt an individual person must have certain beliefs about his or her situation and perhaps some dispositions to act in certain ways. Gilbert's analysis of shared feelings of guilt appears to be an analysis of shared belief rather than an analysis of shared emotion. According to Gilbert, since it is difficult to

distinguish feeling guilt from judging that one is liable for a wrong, it is clear that feeling sensations are not supposed to account for the difference. If we accept this kind of cognitivism, the problem is how to explain the case in which we can see someone is having an irrational feeling: an anti-war protestor, for example, who does not jointly *agree* to the Iraq-war, but nevertheless feels guilty about it. As a citizen of the United States someone may feel she is party to a joint commitment, and in virtue of this she shares responsibility for—every harm her government commits. This might be the case even if she strongly objects to their actions. Although this is true, it seems odd to say that it would be appropriate for an anti-war protestor to feel guilt in response to the Iraqi war. Before discussing the appropriateness of irrational guilt, let us look at how Gilbert could explain irrational collective guilt.

Irrational collective guilt is possible, Gilbert might say, when a group adopts a collective belief and each individual member of the group may express or act on a belief that is not her own. In this respect, Gilbert stresses that one can even become part of a plural subject without entering into an explicit agreement to that effect. Moreover, she adds, "nor need they ever explicitly acknowledge that a certain view is the group's view" (Gilbert 1989, 293). What then is the truth condition of the ascription of group belief? According to Gilbert, it is 'letting a certain view stand as the view of the group' in terms of a 'joint commitment to accept as a body' (Gilbert 1996, 7–15). Hence, Gilbert would say that collective feelings of guilt are a function of the 'joint commitment' to form a unified subject that expresses guilt. But how could there be a unified subject that feels guilt? Let's consider Gilbert's answer to this question.

According to Gilbert, in order to give a proper account of collective guilt, it is inadequate to take the model of individual guilt feelings (Gilbert 2000, 120). Instead, she suggests that the guilty feelings of a group are explained in terms of 'feeling guilt as a body.' How then does a collection of individual's guilt become a collective guilt? On Gilbert's account, it is possible in virtue of being 'the plural subject of a feeling of guilt' which can be defined as individuals being 'jointly committed to feeling guilt as a body.' Gilbert calls this type of joint commitment an 'authority-producing' joint commitment. It is a type of commitment made by the members of the

collective, as a body, through the means of a representative, by authorizing 'some person or body to make decisions' for the collective (Gilbert 2000, 127).

Given this, as Mikko Salmela points out, "it is one thing to make a commitment, either individual or joint, to feel an emotion and another to hope that the feeling emerges, for we cannot make ourselves feel at will" (Salmela 2012, 36). However, "there is no direct way to summon an emotion by committing oneself to feeling it," while, "we indirectly commit ourselves to emotions by jointly committing ourselves to goals and other concerns" (Salmela 2012, 36). As Salmela clarifies, "such commitment rationally commits us to different feelings for goals depending on the outcome—joy if the goal is reached, fear if our progress toward the goal is threatened, disappointment if we fail to reach the goal, and so on" (Salmela 2012, 36).

Now let us consider how we indirectly commit ourselves to feeling emotions. Gilbert claims that in order for a member to participate in plural subjecthood, he or she should 'participate in believing that p as a body' (Gilbert 1994, 251). Hence, Gilbert argues that collective beliefs provide 'individuals with a sense of unity or community with others' (Gilbert 1994, 253).

Given Gilbert's account, in order to explain the case in which a person feels collective guilt although they played no part in the harm committed by the group, we should make a distinction between 'believing that p…' and 'accepting that p.' This is because one can accept something even when they feel it to be false; whereas one cannot *believe* in such a falsity. As K. Brad Wray points out, the views as adopted by plural subjects are a means to realizing the group's goals, however, belief that reflects truth or falsity (or accuracy) is not concerned with those goals (Wray 2001, 324). In order to understand this, let us consider Gilbert's example. Consider when two parents decide that their child should be home at a certain time in the evening. It is the parents as a unit that believe this, despite the fact that neither parent may individually believe that this is when their child should be home (Gilbert 1994, 249–50). The normativity the parents could attach to the 'should be home' is: one must believe (truly or falsely) that the others in the group individually accept the collective item. The parents are jointly committed to raising their child properly. Hence, as a unit, the parents develop views with this goal in mind. It is their collective goal that determines what they claim to 'collectively believe.' Sometimes they will even

adopt views that conflict with their personal preferences and beliefs. If this is right, the relevant distinction Gilbert should make is a distinction between my believing that p and my accepting (and believing) that we collectively believe that p.

Even if we can clarify the case that a person feels collective guilt although they played no part in the harm committed by the group, by distinguishing between accepting that p… and believing that p…, another problem is how to explain the *appropriateness* of the case of irrational guilt. I shall try to answer this in section 5 of this paper. Before looking at this problem, let us consider the difference between guilt and shame.

One might say that we commonly treat the terms 'shame' and 'guilt' interchangeably. However, the difference between guilt and shame lies in their functions. Although guilt may feel similar phenomenologically to shame, it functions quite distinct from shame in that the latter is usually directed at one's *self* rather than one's *actions*. It is sometimes suggested that guilt is act-directed while shame is person-directed. In this respect, Jesse Prinz observes that "[A] guilty person can feel that her actions were wrong without feeling like a bad person" (Prinz 2007, 77). If you are a particularly shame-prone person, you can feel shame about not only the failure of your conduct but also the intention to do so (Adam Morton 2013, 180–82). This is because "shame essentially relates to our 'social selves' i.e., those properties related to the standing we have in the eyes of others" (Deonna and Teroni 2011, 195). But guilt is what you should feel about the actual failure.

If we agree with the view that guilt is act-directed whereas shame is person directed, we can say that the proper response to someone's blames for your alleged failing you do not think you did is *anger* or *indignation* rather than *shame*. If this is true, it makes difficult to explain our case, i.e., anti-war protestor's vicarious guilt. The reason is that if the guilt is not tied to *the self* but an emotion related to *the act*, the anti-war protestor does not need to feel guilty about the actions that she did not do. In order to solve this problem, I suggest a way to feel guilty on behalf of others even if I did not do the wrong action. I suggest that we should take feelings of guilt as a *reactive attitude*: one that we have in reaction to the actions of *ourselves* and *others*.

Given that guilt functions quite distinct from shame in that the one is usually directed at one's *action* rather than one's *self,* one might argue that feelings of guilt can never be felt vicariously because guilt is tied to what oneself *does*. Yet, one might argue against this argument: someone can have vicarious feelings of guilt by identifying with you as her (by assimilation or identification). If this this true, then why can't she feel just about anything you feel? The reason is that if we endorse the cognitivism of emotion, we cannot explain 'somatic feelings' when we have the emotions, because cognitivists deny that emotions are constituted both by judgment/belief and by somatic feelings.[2] Hence, I suggest that we should take reactive attitudes in order to explain vicarious feeling. As P. F. Strawson describes it, "reactive attitudes are essentially natural human reactions to the good will or ill will or indifference" people have towards each other (Strawson 1974, 10).

Emotions such as guilt, resentment, and indignation are what Strawson calls reactive attitudes. They provide the key to understanding moral responsibility and its conditions. To say that someone is responsible is to say that she is just prone to have these attitudes towards others and to be responsible is just to be the appropriate target of these attitudes. Strawson classifies reactive attitudes into three categories: personal, vicarious, and self-reactive, but he argues that moral reactive attitude should be an attitude that is felt in place of others, such as moral indignation or disapproval, rather than an individual attitude. Such attitudes are the criteria for actions and attitudes to be taken about others, not about oneself (Strawson 1974, 70–71). Given this, taking feelings of guilt as a reactive attitude will help us justify indirectly attributing responsibility by way of directly attributing the property to its members.

3. Irrational guilt and recalcitrant emotion

Having established that the proper response to group harms is a reactive attitude, let us distinguish, following Jesse Prinz (2007), 'reactive' from

[2] Among the cognitivists, Solomon(1993) and Nussbaum (2001) are proponents of this view. Gilbert seems to support this position.

'reflective' emotions. These correspond respectively to what Aron Ben-Ze'ev calls 'other-blaming' and 'self-blaming' emotions (Ben-Ze'ev 2000). We can classify the reactive attitudes into two categories: the vicarious, and the self-reactive. The vicarious reactive attitudes are those attitudes we feel in response to ill or good will shown to others (e.g. disapprobation and indignation or approval and support). The self-reactive attitudes are attitudes directed at ourselves in response to how we treat other and ourselves. Guilt, for example, is a response to our own actions. While these two kinds of moral emotions are negative emotions (i.e., indignation and guilt), as Prinz points out, there are also positive moral emotions, such as sympathy. Hence, Prinz suggests, there is a general asymmetry between positive and negative emotions in morality: "desirable behavior is more likely to be shaped through negative emotions than positive" (Prinz 2007, 79). This is because societies have a greater interest in eliminating bad behaviour than in promoting especially good behavior. In order to get rid of bad behavior, punishment can be more effective than praise. Assuming that punishment produces negative emotions, it can be said that we follow moral rules in order to avoid the emotional price of bad behavior. The emotional price of bad behavior would be guilt in the case of self-blame and anger in the case of other blame.

If we accept the idea that guilt is constitutively tied to personal responsibility, in the case of the anti-war protestor discussed the above (in section 2), she does not feel guilty about the actions of her government, but guilty about her own failure to prevent it. On the other hand, we can say that although the individual thinks that she is herself unlikely to participate in such collective acts or even to allow them, it might be enough just to feel outrage, or indignation in response to the actions of her government.

Given my suggestion that the proper response to group harms is 'anger' or 'indignation,' one might call into question where the normative dimension enters into those emotions.[3] Before answering this question, let us consider Allan Gibbard's reactive emotion view.

[3] In light of this, Gibbard's view has been criticized for failing to distinguish emotions such as anger from the normative dimension.

Gibbard is worried about the judgmentalist's accounts because he wishes to understand moral evaluations of wrongness and blameworthiness in terms of reactive emotions. According to him, to think and to act morally or blameworthily "is to accept norms that prescribe, in such a situation, guilt on the part of the agent and resentment on the part of others" (Gibbard 1990, 47). He argues that the judgmentalist cannot allow for inappropriate guilt and resentment (Gibbard 1990, 130, 147, 148–49). Although he does not address the problem of collective guilt, he would agree that the anti-war protestor's guilt in our example is irrational and cannot be explained in terms of judgmentalism. We may call this kind of irrational emotion 'recalcitrant emotion.'[4] If I believe that a rabbit is not harmful, according to cognitivists, then I should not fear it. If I do, it follows that it is a mistake to fear the rabbit, and the fear should be abandoned. But it is possible that fear persists, even when I consciously believe that the rabbit is harmless. The recalcitrant emotions of groups, as Susan James points out, are even harder to dismiss. For example, as James puts it, a member of the American Republican party may fear Islamic fundamentalism, and hold the unchanging belief that fundamentalists are dangerous, despite evidence to the contrary (Susan James 2003, 228). If the strong cognitivist view were true, emotional recalcitrance would seem to predict that people can have inconsistent beliefs. Hence it follows that emotional recalcitrance gets strong cognitivism into trouble because they take the object of an emotion to be its propositional content. If an intentional object of emotion is one that has a propositional content, cognitivists might run into a contradictory state, because one cannot rationally assert both that p and that $\neg p$ at the same time: both cannot be true at the same time. In this respect, many contemporary emotion theorists suggest that emotions that are conflicting with judgment can be called 'recalcitrant emotion' and compare the recalcitrant mental state with optic illusions of the Müller-Lyer lines: the Müller-Lyer lines continue to appear to be of different lengths while they are known to be equal lengths.[5]

[4] See (Yang 2009a) for a more detailed discussion of recalcitrant emotion.

[5] For example, (D'arms and Jacobson 2003; Susan James 2003). C. Tappolet in her recent work (2012) argues that recalcitrant emotions are a kind of perceptual illusion.

4. Appropriateness of reactive attitude

It would seem appropriate for an anti-war protestor to feel indignation at the War because that response is a manifestation of a sentiment as well as an endorsement of that person's judgment that the government has violated the demands of justice, a demand to which the protestor believes it ought to be held accountable. Indignation is a reaction of public outrage at social injustice, thereby producing resistance and insurrection. This emotion can be categorized as a collective emotion, along with guilt and shame.

This idea supports Gibbard's view, which assimilates reactive emotions closely to other emotional and adaptive states. On Gibbard's (1990) norm expressivism, moral judgments of right and wrong are connected to the rationality of what are identified as the moral emotions—guilt and impartial anger. The general strategy is to treat resentment or anger as identified by their characteristic causes and by the forms of expression and behavior to which they characteristically give rise, and to treat guilt as a refinement of a more basic biological adaptation in specific cultural circumstances.

As we have seen, Gilbert's judgmentalism does not allow us to ascribe those emotions to people who find the emotions uncalled for from their accepted perspective of moral demand. She follows strong cognitivism in rendering emotions as evaluative judgments and feelings as sensations that have only a contingent role in emotion. But in seeking an alternative to this kind of judgmentalist account, I endorse Gibbard's view, which goes to the opposite extreme, denying altogether the role of beliefs in explaining the reactive attitudes.

According to Gibbard, the adaptive function of guilt remains constant across the species, and its function can be explained only in terms of its promoting sensitivity to others' anger. Now, if our feelings can fall under moral evaluations, how can his theory handle questions about the rightness of feeling guilt? There are two questions here: one about the feeling of guilt, and another about its being warranted. Gibbard suggests that to call a feeling warranted is to express one's acceptance of norms that allow or require having that feeling.[6] What then does 'norm' mean for this account? In order

[6] See (Yang 2009b) for a more detailed treatment of this problem.

to answer this question, Gibbard presents a notion of wrongness or rightness as being to some degree culturally specific:

> What a person does is morally wrong if and only if it is rational for him to feel guilty for doing it, and for others to resent him (Gibbard 1990, 42).

Gibbard uses what is rational and what makes sense interchangeably. It is important to note that *what is rational* and *what makes sense* are not meant to be understood in any technical sense. Instead, these terms should be understood as work-a-day normative terms. I follow him in this. Gibbard's sense of 'rational' implies that guilt involves endorsement of a norm requiring. He suggests that norms be understood in terms of the following question:

> Are there situations in which, no matter what the agent does, it will make sense for him to feel guilty for having done it, and for others to be angry at him for having done it? (Gibbard 1990, 43)

So, in the case of guilt, it can be said that guilt is a mechanism that was an evolutionary adaptive response to anger. It makes groups of society reconcilable for mutual cooperation rather than conflict. Animals show a limited response to hostility, while humans need to be able to alleviate anger to bring about reconciliation. Guilt has evolved to regulate these responses. The problem with this explanation is that it is not clear why guilt and anger should be reciprocal. Gibbard seems to see that the standard of rightness or wrongness for feeling guilt is a response-dependent one. This is so because we feel guilty in situations where, in many cases, it is reasonably justified that others are angry with us. According to Gibbard, guilt and anger are supposed to be mutual feelings in that "guilt aims to placate anger, and it is governed by the same norms as govern anger" (Gibbard 1990, 139). Denying this reciprocity makes the Gibbard's adaptive syndrome theory difficult. I am endorsing it.

Gibbard, as we have seen, focuses on the reactive attitudes we have toward individuals and ourselves. But we also have reactive attitudes toward collectives. We can say,

> A group X is morally blameworthy for doing action F in circumstance C just in case it is rational (it 'makes sense') for both

(i) members of group X to feel guilt, and
(ii) others to resent the group X.

There are normative standards that apply not to individuals but to groups, governments, and so on. Consider the case of the terrorist organization Islamic State (IS), for example.

At this point one might raise a question: in order for collectives to be appropriate targets of our moral sentiments, must collectives themselves be able to have reactive attitudes? Following Deborah Perron Tollefsen, I admit the possibility that there are collective reactive attitudes: "Collective guilt may involve an attitude in response to a collective action done by the group of which one is a member" (Tollefsen 2003, 220). I argue that if a group is responsible for an evil act, they are justifiably liable to the anger and resentment of others, and if such attitudes are present in our interactions with collectives, we can attribute moral responsibility to collectives. If we assume that our reactive attitudes toward collectives are tracking the same features as they track at the individual level, as Tollefsen points out, then a further question arises: how then could a collective itself have reactive attitudes?[7] If these attitudes are emotions and emotions are constituted both by judgement/belief and by somatic feelings, one might ask: how could the collective itself *feel*?[8]

If we endorse the strong cognitivism on emotion, following Nussbaum (2001), it is difficult to explain the affective aspect of emotion. This is because, as we have seen in section 1, for Nussbaum an emotion can be analyzed solely in terms of beliefs, or evaluative judgments. Moreover,

[7] This is because, for Strawson the reactive attitudes are reciprocal. Tollefsen mentions that this point made by M. Bratman. See (Tollefsen 2003, 231).

[8] When I say that emotions are feelings, I mean 'somatic feelings,' following the William Jamesian, according to which emotions can be identified with bodily sensations that have a certain pattern. Yet if emotions were merely perceptions of the body, they would represent body as being in such and such a state. This theory also encounters difficulty in explaining the intentionality of emotion, for as many point out, feeling theorist cannot explain the fact that emotions have intentional content. William James sometimes highlighted the turbulence of emotion rather than their intentionality. Hence, for him to experience emotion is to be in some state of agitation, commotion, excitation, etc.

Nussbaum believes that physiological changes are just a *symptom* of a given emotion, but are not necessary for the concept of emotion. However, if this kind of cognitivism were correct, then the cognitivist account of emotion would render the component of feeling in constituting emotion superfluous. I reject this kind of cognitivism on the ground that it ignores a 'feeling component.' *Pace* Nussbaum, if affect is not contrasted with cognition, then the real problem of explaining emotion is how we can develop a thorough account of how emotions dissolve the distinction between thoughts and feelings, or cognition and affect.[9]

Gilbert seems to endorse the strong cognitive theory of emotion, following Nussbaum, when she addresses the collective remorse.

Consider Gilbert's account of collective remorse again:

> Group G feels remorse over an act A if and only if the members of G are jointly committed to feeling remorse as a body over act A (Gilbert 2000, 135).

According to Gilbert, group guilt is a function that is bound together to form an integrative subject that expresses guilt. Individuals who see themselves as members of a group will do what they can, through actions or utterances, to form such subjects.

Yet the difficulty that Gilbert faces is to explain the phenomenological features of such a feeling. Since the group lacks a natural body, it is difficult to say that it is the subject of phenomenological feelings such as 'pangs' or 'twinges' of guilt. Yet, it seems to me that this is not an insuperable problem. For it seems that to hold a corporation or other collective morally responsible is simply to be prone to having the appropriate reactive attitudes toward them and for a collective to be responsible is for them to be subject to the appropriate attitudes. This is because, as many point out, our reactive attitudes are emotional and adaptive states. Emotional and adaptive states, disgust and sadness for example, are the basic emotions that have the phenomenal features. "Indignation may be anger calibrated to injustice, and guilt may be sadness calibrated to self-caused harm" (Prinz 2007, 67).

[9] A more detailed discussion on how emotions dissolve the distinction between thoughts and feelings, or cognition and affect is beyond the scope of this paper.

Anger, for example, is the basic emotion which has the phenomenal feature of being a reaction to and appraisal of a threat or offense. The reason why we can say that anger is an adaptive state is, as Prinz remarks, that "[W]hen we react emotionally to victimization, the anger response is natural because it is evolved to cope with threats, and it disposes us to aggression" (Prinz 2007, 70). Given this, we need to consider whether reactive attitudes are constitutive of the practice of moral responsibility for collectives.

If reactive attitudes are constitutive of moral responsibility, we can say that the proper response to group wrongdoing should be mutually expressed emotions—guilt and anger or indignation, which are self-reactive and vicarious reactive attitudes, respectively. In addition, following Tollefsen (2003), we express our indignation and resentment on collectives themselves.

5. Conclusion

Hence, in this paper, I classify the reactive attitudes into three categories: self-reactive, vicarious, and collective reactive. The self-reactive attitudes are attitudes directed at ourselves in response to how we treat others and ourselves. Guilt, for example, is a response to our own actions. Guilt is a self-reactive attitude directed at ourselves in response to how we treat others and ourselves. In the case of the anti-war protestor, we can say that she feels guilt over the actions of her government, which she does not support, because, being a member of the collective to which the government also belongs (e.g., the United States), she has indirectly accepted those actions by the government. Additionally, she might have a vicarious feeling, i.e., outrage or indignation. In this case, it is a vicarious reactive attitude, on behalf of those victims. As Tollefsen points out, "our indignation on behalf of the victims is directed not just at the individuals who committed and participated in the cover-up, but also at the institution which concerned it and, in certain cases, made it possible" (Tollefsen 2003, 224).

Let us consider how the vicarious reactive attitude, that is, indignation, is possible. In the case of the terrorist attack of 9/11, whether as a citizen of the U.S or from other country, someone might feel outrage or indignation

on behalf of the victims or their families. This is the vicarious reactive attitude. This is the attitude we feel in response to ill or good will shown to others. Indignation is anger calibrated to injustice. One might object that the vicarious reactive attitude of indignation is not linked to phenomenal feeling. Yet, we can say that it can be linked to phenomenal feeling by virtue of the fact that it derives from the basic form of anger, which is concerned with an appraisal of threat or offense. Although righteous anger and indignation have a common ground in that they derive from a more basic form, Prinz differentiates them in that the latter, but not necessarily the former, always involves violations of justice.[10] To conclude, we can say that the appropriateness of the vicarious feeling of group harm derives from a reflection on the appropriateness of our own reactive attitude, that is, vicarious reactive attitude, e.g., indignation or outrage. Hence, in our case, i.e., the anti-war protestor case, it can be said that it would be rational for her to feel indignation for the Iraqi war because she believes her government has violated the demands of justice, a demand to which she believes it ought to be held.

Acknowledgements

Previous versions of some parts of this paper were presented at the 7th International Conference on Collective Intentionality at the University of Basel 23-26 August 2010 in Switzerland and at the 9th International Conference on Collective Intentionality at Indiana University in Bloomington 16-19 September 2014. I am very grateful to those audiences for the valuable discussions. Especially, I would like to thank Kirk Ludwig, who gave excellent comments on early version of it. I also thank Christopher Kutz and Ronald de Sousa, who suggested some important improvement in personal correspondence.

Funding

This work was supported by Hankuk University of Foreign Studies Research Fund of 2020.

[10] For a more detailed discussion on anger and indignation, see (Prinz 2007); see also (Yang 2009b).

References

Ben-ze'ev, Aaron. 2000. *The Subtlety of Emotions,* Cambridge, MA: MIT Press.

D'Arms, Justin and Jacobson, Daniel. 2003. "The Significance of Recalcitrant Emotion." In *The Philosophy and the Emotion*, edited by Anthony Hatzimoysis, 127–46. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511550270.009

Deonna, Julien and Teroni, Fabrice. 2011. "Is Shame a Social Emotion?" In *Self-Evaluation: Affective and Social Grounds of Intentionality*, edited by Antica Konzelmann Ziv, Keith Lehrer, and Hans Bernhard Schmid. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-1266-9_11

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings.* Cambridge, Mass.: Harvard University Press.

Gilbert, Margaret. 1989. *On Social Facts.* New York: Routledge

Gilbert, Margaret. 1994. "Remarks on Collective Belief." In *Socializing Epistemology*, edited by Frederick Schmitt, 235–56. Maryland: Rowman & Littlefield.

Gilbert, Margaret. 1996. *Living Together.* Maryland: Rowman& Littlefield

Gilbert, Margaret. 1997. "Group Wrongs and Guilt Feelings." *The Journal of Ethics* 1 (1): 65–84. https://doi.org/10.1023/A:1009712003678

Gilbert, Margaret. 2000. *Sociality and Responsibility.* Blue Ridge Summit: Rowman and Littlefield.

Gilbert, Margaret. 2002. "Collective Guilt and Collective Guilt Feelings." *The Journal of Ethics* 6: 115–43. https://doi.org/10.1023/A:1015819615983

James, Susan. 2003. "Passion and Politics." In *Philosophy and the Emotions*, edited by Anthony Hatzimoysis, 221–34. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511550270.014

Morton, Adam. 2013. *Emotion and Imagination.* Cambridge: Polity Press.

Nussbaum, Martha. C. 2001. *Upheavals of Thought.* Cambridge: Cambridge University Press.

Prinz, Jesse. 2007. *The Emotional Constructions of Morals.* Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199571543.001.0001

Salmela, Mikko. 2012. "Shared Emotion." *Philosophical Explorations* 15 (1): 33–46. https://doi.org/10.1080/13869795.2012.647355

Solomon, Robert C. 1993. *The Passions: Emotions and the Meaning of Life.* Second edition. Indianapolis, IN: Hackett.

Strawson, Peter F. 1974. "Freedom and Resentment." Reprinted in *Free Will*, edited by Gary Watson, 72–93. Oxford: Oxford University Press.

Tappolet, Christine. 2012. "Emotions, Perceptions, and Emotional Illusions." In *Perceptual Illusions: Philosophical and Psychological Essays*, edited by Clotilde Calabi, 205–22. London: Palgrave Macmillan.

Taylor, Gabriele. 1985. *Pride, Shame, and Guilt: Emotions of Self-Assessment*. Oxford: Oxford University Press.

Tollefsen, Deborah. 2003. "Participant Reactive Attitudes and Collective Responsibility." Philosophical Explorations 6 (3): 218–34. https://doi.org/10.1080/10002003098538751

Yang, Sunny. 2009a. "Emotion, Intentionality and Appropriateness of Emotion." *Organon F* 16 (1): 82–104.

Yang, Sunny. 2009b. "Appropriateness of Moral Emotion and a Humean Sentimentalism." *The Journal of Value Inquiry* 43: 67–81. https://doi.org/10.1007/s10790-009-9144-1

Yang, Sunny. 2016. "Do Emotions Have Directions of Fit?" *Organon F* 23 (1): 32–49.

Wray, K. Brad. 2001. "Collective Belief and Acceptance." *Synthese* 129 (3): 319–33. https://doi.org/10.1023/A:1013148515033

Organon F 27 (3) 2020: 395–410
https://doi.org/10.31577/orgf.2020.27306

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Does the Conceivability of Zombies Entail Their Possibility?

Karol Polcyn*

Received: 11 July 2019 / Accepted: 10 February 2020

Abstract: According to the two-dimensional argument against materialism, developed by David Chalmers, the conceivability of zombies entails primary possibility, and the primary possibility of zombies entails further secondary possibility. I argue that the move from the conceivability to primary possibility of zombies is unjustified. Zombies are primarily impossible despite being conceivable if the corresponding phenomenal and microphysical concepts have coinciding primary intensions (refer to the same properties in all possible worlds considered as actual) despite being distinct concepts. But there is no good reason to think that phenomenal and microphysical concepts cannot have coinciding primary intensions despite being distinct concepts. As I argue, this conclusion follows from reflection on special cognitive features of phenomenal concepts.

Keywords: Conceivability; consciousness; phenomenal concepts; possibility; two-dimensional semantics.

1. Introduction

It seems that we can conceive of zombies: beings identical to us physically and functionally but lacking phenomenally conscious states. We are

* University of Szczecin
Institute of Philosophy and Cognitive Science, University of Szczecin, ul. Krakowska 71-79, 71-017 Szczecin, Poland
karol.polcyn@gmail.com

phenomenally conscious in the sense that there is something *it is like* to undergo various experiences that we undergo, such as seeing red, feeling pain, being surprised, etc. By contrast, zombies have no such inner life. There is nothing *it is like* to be a zombie.

Assuming that zombies are conceivable, one crucial question is whether the conceivability of zombies entails that zombies are possible. Materialism is the view that zombies are impossible, which amounts to the view that consciousness is necessitated by physical properties. If the conceivability of zombies entails possibility, the conceivability of zombies entails that materialism is false. This issue has been central to the recent debate on the nature of consciousness.

The most influential argument to the effect that the conceivability of zombies entails possibility has been developed by David Chalmers. Chalmers (1996, 2010) articulates his argument in terms of two-dimensional semantics, by distinguishing two senses of the possibility of statements: S is *primarily possible* if S is true at a possible world considered as actual, and S is *secondarily possible* if S is true at a possible world considered as counterfactual. Chalmers argues that the conceivability of zombies entails that zombies are primarily possible. From this he infers further (under some qualifications to be clarified shortly) that zombies are secondarily possible.

Here I argue that the two-dimensional argument is far from successful. The argument seems to fail in its crucial step from the conceivability to primary possibility of zombies. In effect, the view that zombies are secondarily impossible despite being conceivable—in Chalmers' (2002a) classification, type-B materialism—remains unthreatened.

2. The two-dimensional argument

Assume that P is a conjunction of all microphysical truths (including microphysical laws) and that Q is the truth that someone is conscious. Then the claim that a zombie world is conceivable is the claim that $P\&{\sim}Q$ is conceivable. $P\&{\sim}Q$ says that everything is microphysically as in our world but that no one is conscious. In this sense, $P\&{\sim}Q$ says that the world is a zombie world.

It is fair to assume that $P\&{\sim}Q$ is conceivable: $P\&{\sim}Q$ expresses a coherent hypothesis, that is, a hypothesis that cannot be ruled out a priori. The hypothesis is coherent, because there are no conceptual links between microphysical concepts, characterizing the world in terms of structure and dynamics, and phenomenal concepts, that is, the concepts of the phenomenal characters of experiences.[1,2]

Chalmers argues further that the conceivability of $P\&{\sim}Q$ entails that $P\&{\sim}Q$ is possible. At the most general level, the structure of his argument is as follows. The conceivability of $P\&{\sim}Q$ entails that $P\&{\sim}Q$ is primarily possible. But the primary possibility of $P\&{\sim}Q$ entails further—under certain assumptions about the nature of fundamental physical properties—that $P\&{\sim}Q$ is secondarily possible.

Chalmers' argument has shaped the debate on the nature of consciousness in the last two decades, and it continues to be puzzling to materialists. Is it right to think that the conceivability of $P\&{\sim}Q$ entails that $P\&{\sim}Q$ is primarily possible? Before evaluating the argument, let me present it in a more explicit way. We begin with a more explicit articulation of the key distinction between primary possibility and secondary possibility.

Intensions for statements are functions from possible worlds to truth values; correspondingly, intensions for concepts are functions from possible worlds to extensions. The *primary intension* of S is true at a possible world w (w verifies S, or S is true at w considered as actual) if and only if the following condition is met: if we came to accept that our world is qualitatively like w, we should endorse S.[3] On the other hand, to say that the *secondary intension* of S is true at w is to say that w satisfies S in the sense that S is true at w considered as counterfactual. Correspondingly, S is *primarily possible* (1-possible) if the primary intension of S is true at some centered world, and S is *secondarily possible* (2-possible) if the secondary

[1] Following Chalmers (2002b), we can assume that $P\&{\sim}Q$ is ideally (as opposed to *prima facie*) conceivable: $P\&{\sim}Q$ cannot be ruled out a priori even on ideal rational reflection, that is, even when we abstract away from our cognitive limitations.

[2] For a comprehensive defense of the conceivability of zombies against potential objections, see (Chalmers 2002a, 2010, sec. 4 and sec. 5).

[3] Possible worlds considered as actual are *centered* worlds: worlds marked with a specified individual and time.

intension of S is true at some uncentered possible world. According to this framework, the negations of Kripkean a posteriori necessities are 2-impossible but not 1-impossible. For example, 'Water is not H_2O' is 2-impossible, but it is not 1-impossible. The primary intension of 'Water is not H_2O' is true at the centered possible world with XYZ in the oceans and lakes: if we came to accept that our world is the XYZ world, we should accept 'Water is not H_2O' (Chalmers 1996, ch. 2 and ch. 4, 2002b, 2006, 2010; Chalmers and Jackson 2001).

Given the above framework, Chalmers argues as follows. There is no gap between ideal conceivability and 1-possibility. In other words, ideal conceivability entails 1-possibility. If so, the ideal conceivability of $P\&{\sim}Q$ entails that $P\&{\sim}Q$ is 1-possible. If P has the same primary and secondary intensions and Q has the same primary and secondary intensions, it will follow that $P\&{\sim}Q$ is 2-possible. But here one must be careful. It is uncontroversial that Q has the same primary and secondary intensions, because the concept of consciousness has the same primary and secondary intensions: the primary intension of the concept of consciousness picks out a phenomenal feel, and the secondary intension picks out a phenomenal feel as well.[4] On the other hand, it is plausible to hold that the primary and secondary intensions of microphysical terms do not coincide—this will be the case if microphysical terms pick out the intrinsic properties that serve as the categorical bases of microphysical dispositions (Chalmers 2010, sec. 3). If one accepts this view, the primary and secondary intensions of P will not coincide. And if the categorical bases of microphysical dispositions are phenomenal or protophenomenal properties, $P\&{\sim}Q$ will be 1-possible but 2-impossible. However, the view that the categorical bases of microphysical dispositions are phenomenal or protophenomenal properties (Russellian monism) can be left aside, given its highly speculative status. So, when this view is left aside, the 1-possibility of $P\&{\sim}Q$ entails 2-possibility. That is, when Russellian monism is left aside, the entailment from the 1-possibility to 2-possibility of $P\&{\sim}Q$ is justified even if the primary and secondary intensions of physical concepts do not coincide (even if P has different primary and secondary

[4] If something feels like a conscious experience, even in some counterfactual world, it *is* a conscious experience (Chalmers 1996, 133). By contrast, a substance that looks like water in a counterfactual world but is not H_2O is not water.

intensions). In the light of the above analysis, the two-dimensional argument can be articulated as follows (Chalmers 2010):

1. $P\&{\sim}Q$ is conceivable.
2. If $P\&{\sim}Q$ is conceivable, $P\&{\sim}Q$ is 1-possible.
3. If $P\&{\sim}Q$ is 1-possible, $P\&{\sim}Q$ is 2-possible.
4. If $P\&{\sim}Q$ is 2-possible, materialism is false.
5. So, materialism is false.

A note on the intensions of microphysical concepts. As Chalmers (2002a, sec. 11) points out, whether microphysical terms have different primary and secondary intensions, namely, whether those terms pick out intrinsic properties rather than microphysical dispositions is a terminological rather than a substantive issue. If so, we will stipulate here that the primary and secondary intensions of microphysical concepts coincide. This will simplify the subsequent discussion.

Now, how does Chalmers justify premise 2? Premise 2 follows from the general principle that if S is ideally conceivable, S is 1-possible. This principle (CP) comes down to the point that there is no gap between conceivability and 1-possibility. Chalmers (2010, sec. 8) points out that there are no clear counterexamples to CP. In particular, the principle is consistent with Kripkean a posteriori necessities, as we saw earlier. But why suppose that CP must be true? To justify this, Chalmers argues that the view that there is a gap between conceivability and 1-possibility amounts to the unacceptable view that a space of metaphysically possible worlds is narrower than the space of ideally conceivable worlds. Chalmers' reasoning goes here as follows. For any sentence S, if S cannot be ruled out a priori, there is a *scenario* (a maximal, a priori coherent hypothesis about the character of the actual world) that verifies S in the sense that it would be incoherent to suppose that the scenario obtains and S is not true (Chalmers 2002b, 2006, 2010, sec. 7). For example, the scenario verifying 'Water is not H_2O' involves the assumption that the transparent liquid that fills the oceans and lakes is not H_2O but, say, XYZ: if we suppose that this scenario actually obtains, we should accept 'Water is not H_2O.' If S is conceivable and 1-possible, there is a possible world that corresponds to the scenario that verifies S. The possible world at issue is the world that verifies S. But if S is both conceivable and 1-impossible, there is no possible world corresponding

to the scenario that verifies S (Chalmers 2010, sec. 7). Chalmers (2010, sec. 10) argues then that there cannot be scenarios that correspond to no possible worlds. As he points out, we have no concept of independent metaphysical modality, introducing a separate space of metaphysically possible worlds, in addition to logically possible worlds. Independent metaphysical modality plays no role in our conceptual system.

3. The unjustified move from the conceivability to primary possibility of zombies

In reply to the two-dimensional argument, I take it that premise 2 is unjustified. Leaving the details of how Chalmers justifies this premise aside, it is arguable that there is no good reason to think that $P\&{\sim}Q$ cannot be 1-impossible despite being conceivable. We can see this if we articulate the content of the claim that $P\&{\sim}Q$ is 1-impossible despite being conceivable in terms of the intensions of phenomenal and microphysical concepts.

What are the conditions under which it is true that $P\&{\sim}Q$ is 1-impossible despite being conceivable? Well, $P\&{\sim}Q$ is 1-impossible despite being conceivable if the corresponding phenomenal and microphysical concepts have coinciding primary intensions despite being distinct concepts, that is, if they refer to the same properties in all possible worlds considered as actual. This can be analyzed further. To say that two distinct concepts have coinciding primary intensions is to say that two distinct concepts (i) corefer and (ii) are *epistemically rigid* in the sense that each concept picks out the same thing in all possible worlds considered as actual. If so, we can say that $P\&{\sim}Q$ is 1-impossible despite being conceivable if the corresponding phenomenal and microphysical concepts corefer, are distinct and epistemically rigid.

Now, phenomenal and microphysical concepts are, by assumption, a priori distinct. They are also epistemically rigid. As Chalmers (2003) points out, each phenomenal concept picks out one and the same property in all possible worlds considered as actual. The same can be said about microphysical concepts. The idea that microphysical concepts are epistemically rigid corresponds to our earlier assumption that phenomenal concepts have coinciding primary and secondary intensions. As we saw, this assumption

is not true if we accept Russellian monism. However, here we are leaving Russellian monism aside.

Given that phenomenal and microphysical concepts are a priori distinct and epistemically rigid, the key question is whether they can corefer despite being distinct epistemically rigid concepts. In other words, the question is whether those concepts can refer epistemically rigidly to the same thing despite being distinct concepts. If they can, this will imply that the primary conceivability of $P\&{\sim}Q$ does not entail 1-possibility.

I take it that there are no uncontroversial examples of conceptually distinct epistemically rigid concepts of the same property. Typically, a posteriori true identity claims involve at least one concept that is not epistemically rigid. This is obviously true in the case of Kripkean a posteriori identities, such as 'Water is H_2O' or 'Cicero is Tully.' Materialists responding to the conceivability argument have proposed various examples of identity claims that would break this rule. However, it seems that all such attempts are far from uncontroversial. For example, Kallestrup (2006) considers the identity claim '$Q = H_2O$', where Q is a quantum-mechanical description of H_2O, and argues that this claim is both true a posteriori and 1-necessary. Thus, Kallestrup assumes that Q and H_2O are distinct epistemically rigid concepts of the same thing. In reply, Chalmers (2010, 171–72) argues, however, that Q is not epistemically rigid: it is plausible that something with the structure of Q yields different chemical-level properties due to different quantum-mechanical laws.[5]

Still, the fact that there are no uncontroversial cases of conceptually distinct epistemically rigid concepts of the same thing does not give us the reason to think that there cannot be such concepts. If one thinks that there cannot be such concepts, one would need to show this on independent grounds, by providing some principal reason why conceptually distinct epistemically rigid concepts can never refer to the same thing. Is it plausible to think that there such a reason? In my view, there is no such reason. As I argue below, there is no good reason to think that phenomenal and microphysical concepts cannot be conceptually distinct epistemically rigid

[5] For Chalmers' replies to other proposed examples of strong necessities (statements that are a posteriori and 1-necessary), see (Chalmers 2010, sec. 8).

concepts of the same thing. This conclusion follows from reflection on special cognitive features of phenomenal concepts.

To see one potential reason why phenomenal and microphysical concepts cannot refer to the same thing, compare epistemically rigid concepts with concepts that are rigid without being epistemically rigid, such as the concept *water*. The concept *water* refers to different kinds in different possible worlds considered as actual, which means that we cannot know a priori what it refers to in the actual world. What this concept refers to depends on how the world turns out. Things are different with epistemically rigid concepts, that is, concepts referring to the same property in all possible worlds considered as actual. The reference of such concepts does not depend on how the world turns. Thus, in the case of epistemically rigid concepts, we know a priori what they refer to.

Given that epistemic rigidity has the above implication, it is natural to expect that epistemically rigid concepts referring to the same property will not be a priori distinct. In particular, it might seem that we should expect that if phenomenal and microphysical concepts refer to the same thing, there should be an a priori connection between those concepts. But, of course, it is not true a priori that phenomenal and microphysical concepts corefer. In effect, given our expectation, it would follow that those concepts cannot corefer.

I reply, I do not think that this argument succeeds. Call the expectation that epistemically rigid concepts referring to the same property cannot be a priori distinct 'the expectation of transparency.' I take it that the expectation of transparency is unjustified in the case of the corresponding phenomenal and microphysical concepts. This has to do with the fact that phenomenal and microphysical concepts play very different cognitive roles. Microphysical concepts analyze their referents in theoretical (scientific) terms. But this is not how phenomenal concepts work. Phenomenal concepts do not analyze their referents in theoretical terms. Instead, we use them to refer to our own conscious experiences when we actually undergo those experiences or when we recreate them in imagination.[6] If this is so,

[6] Some philosophers (e.g. Chalmers 2007; Levine 2007) assume that phenomenal concepts *acquaint* us with conscious experiences in the sense that we gain substantive knowledge about our own experiences merely in virtue of undergoing (or imagining)

then we cannot expect to be able to see a priori that phenomenal and microphysical concepts corefer, even if they do as a matter of facts. The different cognitive roles of those concepts will keep them unconnected a priori, even if both concepts refer—epistemically rigidly—to the same thing.[7]

Interestingly, Papineau (2007, 128–32) has another way of explaining why the expectation of transparency in the case of phenomenal and microphysical concepts is unjustified. He argues that it simply does not follow, from the fact that phenomenal concepts are epistemically rigid, that we should be able to see a priori that phenomenal concepts refer to microphysical properties if they do. We can explain why phenomenal concepts are

---

them. However, most of physicalists would not go that far. Instead, they assume simply that phenomenal concepts are exercised in the presence of the experiences being referred to or their imaginative recreations (see Loar 1997; Papineau 2002, 2007). Loar assumes that this is consistent with the idea that phenomenal concepts are type-demonstratives. On the other hand, Papineau assumes that, insofar as phenomenal concepts are exercised in the presence of the experiences being referred to or their imaginative recreations, phenomenal concepts *use* the experiences they refer to (by assumption, imaginative recreations of experiences are phenomenologically similar to actual experiences).

[7] Loar (1997, 1999) argues in a similar way that we cannot expect to be able to see a priori that phenomenal and physical concepts corefer, even though both kinds of concepts refer *directly*, without the need for contingent reference-fixers. This expectation is, according to Loar, unjustified, because phenomenal and physical concepts play different cognitive roles (phenomenal concepts, unlike physical concepts, belong to the class of recognitional concepts, that is, type-demonstratives grounded in our dispositions to classify things by way of perceptual discriminations). I do not think that Loar's account of cognitive differences between phenomenal and physical concepts is correct, since I do not think that phenomenal concepts are demonstratives. Demonstratives have a reference-fixing "character" that leaves their referent open (Kaplan 1989). By contrast, our core phenomenal concepts are tied a priori to specific sorts of qualities (Chalmers 2003). Still, I agree with Loar's general idea that cognitive differences between phenomenal and physical concepts make the expectation of transparency in the case of phenomenal and physical concepts unjustified. The expectation at issue is unjustified when we think of phenomenal and physical concepts as directly referring concepts (which is what Loar argues) but also when we think of phenomenal and physical concepts as epistemically rigid concepts, as I argue here.

epistemically rigid, even if we assume that we do not know a priori that phenomenal concepts refer to microphysical properties. Papineau's explanation turns on the point that phenomenal concepts *use* the experiences they refer to, in the sense that our exercises of phenomenal concepts are typically accompanied by the experiences being referred to or by their imaginative recreations (see note 6). Even though we do not know a priori that phenomenal concepts refer to microphysical properties, argues Papineau, phenomenal concepts are epistemically rigid because of the fact that they use their referents.

In reply, I agree that the use-mention feature of phenomenal concepts explains why phenomenal concepts are epistemically rigid. But I do not see why the use-mention feature of phenomenal concepts should explain why it is consistent to hold both that phenomenal concepts are epistemically rigid and that we do not know a priori that they refer to microphysical properties. After all, one might now ask why phenomenal concepts use their referents, even though we do not know a priori that they refer to microphysical properties. However, it is hard to see what the relevant explanation could be here (Papineau does not provide any such explanation). If so, Papineau's way of responding to the expectation of transparency does not look promising. We can explain why phenomenal concepts are epistemically rigid (in terms of the use-mention feature of phenomenal concepts), but we seem to have no explanation of why it is consistent to hold both that phenomenal concepts are epistemically rigid and that it is not true a priori that phenomenal concepts refer to microphysical properties.

Still, in order to show that the expectation of transparency in the case of phenomenal and microphysical concepts is unjustified, we do not need to explain why phenomenal concepts are epistemically rigid, even though we do not know a priori that they refer to microphysical properties. It is sufficient to appeal to the fact that the corresponding phenomenal and microphysical concepts play different cognitive roles. That is the strategy we have adopted here.

We have explained away one potential objection to the view that phenomenal and microphysical concepts are distinct epistemically rigid concepts of the same thing, namely, the objection based on the expectation of transparency. Is there any room for further objections? One might still

worry that it is arbitrary to suppose that phenomenal and microphysical concepts are distinct epistemically rigid concepts of the same thing. As we have agreed, there are no uncontroversial examples of distinct epistemically rigid concepts of the same thing. If phenomenal and microphysical concepts are examples of such concepts, this example will be unique. But then one might think that postulating such an exception is entirely arbitrary and therefore unacceptable.

I do not think that the above worry is justified. The exceptional status of the phenomenal-physical case is not arbitrary at all. In fact, the reason why this is so is already implicit in what has been said so far. Ask yourself the following question: why cannot we make sense of there being conceptually distinct epistemically rigid concepts of the same thing outside the mind-body domain? Well, we cannot make sense of this because of the expectation of transparency. We expect that nonphenomenal epistemically rigid concepts cannot refer to the same thing unless they are connected a priori. This expectation of transparency in the case of nonphenomenal epistemically rigid concepts is perfectly intelligible. However, the corresponding expectation is not justified in the case of phenomenal and microphysical concepts, given that those concepts play different cognitive roles. So, this explains why it is not arbitrary to suppose that phenomenal and microphysical concepts are unique in being conceptually distinct epistemically rigid concepts of the same thing. The reason why we cannot make sense of uncontroversial examples of distinct epistemically rigid concepts of the same thing outside the mind-body domain simply does not extend to the particular case of phenomenal and microphysical concepts.

To illustrate this, think of the following case. The concept *being a figure with all of its surface points equidistant from its centre* is an epistemically rigid concept of sphericity: it refers to sphericity in all possible worlds considered as actual. Can we think of an epistemically rigid concept of sphericity that would be a priori distinct from the above concept? It seems that we cannot. We expect that any epistemically rigid concept of sphericity should be a priori connected with the concept *a figure with all of its surface points equidistant from its centre*. For example, consider the concept *being disposed to roll when pushed*. This is also an epistemically rigid concept of sphericity. But it is not difficult to see that there is an intelligible connection

between being a figure with all of its surface points equidistant from its centre and being disposed to roll when pushed. In the absence of such a connection, we would have no reason to think that the two concepts at issue are concepts of the same thing.[8] By contrast, we cannot expect that the connection between phenomenal and microphysical concepts should be transparent in the same sense. As we argued above, the radically different cognitive roles of those concepts will keep them unconnected a priori, even if those concepts refer—epistemically rigidly—to the same properties.

Now, one might argue that the case of phenomenal and microphysical concepts is not really different from the case of our two concepts of sphericity. Just as phenomenal and microphysical concepts play different cognitive roles, the concepts of sphericity also play different cognitive roles despite the fact that there is an intelligible connection between them: the former has its place in theoretical mathematical reasoning and the latter plays a role primarily in the practical concerns of everyday life. If this is right, then we haven't made the case that the expectation of transparency is unjustified in the case of phenomenal and microphysical concepts—the fact that phenomenal and microphysical concepts play different cognitive roles does not make this expectation unjustified. We should expect that there is an intelligible connection between phenomenal and microphysical concepts just as much as we expect that there is an intelligible connection between our concepts of sphericity.[9]

In reply, I do not think that the concepts of sphericity in question can be said to play different cognitive roles. I take it that two concepts of the same thing play different cognitive roles if the applications of those concepts put us in different kinds of cognitive relation to the referent.[10] But the kind of cognitive relation we have to sphericity when we use the concept *being a figure with all of its surface points equidistant from its centre* is not

[8] For a similar analysis of this example, see (Goff 2017, sec. 5.6).

[9] I want to thank an anonymous reviewer for raising this point.

[10] This can be illustrated by the concepts *cramp* and *muscle contraction*. Both are concepts of muscle contraction. But the former is a recognitional concept (discriminates muscle contraction perceptually), whereas the latter is a theoretical concept (describes muscle contraction in theoretical terms). (For an analysis of recognitional concepts, see Loar 1997).

different from the kind of cognitive relation we have to sphericity when we use the concept *being disposed to roll when pushed*. After all, both concepts refer to sphericity by way of describing what it takes for an object to be spherical. By contrast, there is a crucial cognitive difference between phenomenal and microphysical concepts: the former, unlike the latter, do not refer to experiences by way of describing them. Instead, they refer to experiences by way of *using* them (in the sense introduced by Papineau and mentioned in note 6). So, there is a clear sense in which the phenomenal-physical case is different from the case of concepts of sphericity. Given the lack of cognitive differences between our concepts of sphericity, we can expect that there should be an intelligible connection between those concepts. By contrast, we cannot expect that there should be an intelligible connection between phenomenal and microphysical concepts.

In the end, the case of phenomenal and microphysical concepts is special due to the unique status of phenomenal concepts among nonphenomenal epistemically rigid concepts in general. Nonphenomenal epistemically rigid concepts typically refer by way of describing their referents (think here of our concepts of sphericity), whereas phenomenal concepts do not refer by way of describing their referents. This explains why the expectation of transparency is unjustified uniquely in the case of phenomenal and microphysical concepts. When we think of nonphenomenal epistemically rigid concepts of the same thing, we expect that such concepts should be connected a priori, since we expect that there should be an a priori connection between epistemically rigid descriptions of the same thing. However, we cannot expect an a priori connection between phenomenal and microphysical concepts, given that phenomenal concepts, unlike physical-theoretical concepts, do not refer via descriptions.

I conclude that there is no good reason to think that phenomenal and microphysical concepts cannot corefer despite being distinct epistemically rigid concepts. If so, there is no good reason to think that the corresponding phenomenal and microphysical concepts cannot have coinciding primary intensions despite being distinct concepts. Consequently, there is no good reason to think that $P\&{\sim}Q$ cannot be 1-impossible despite being conceivable.

One could still raise the following objection to our reply to the two-dimensional argument. We have argued that there is no good reason to

think that the corresponding phenomenal and microphysical concepts cannot refer to the same properties despite being distinct epistemically rigid concepts. But one could argue that those concepts cannot corefer on the grounds that there cannot be the gap between the conceivability and 1-possibility of $P\&{\sim}Q$. By assumption, if the corresponding phenomenal and microphysical concepts refer to the same properties despite being distinct epistemically rigid concepts, phenomenal and microphysical concepts have coinciding primary intensions despite being distinct concepts. This implies further that $P\&{\sim}Q$ is 1-impossible despite being conceivable. Thus, if there is a good reason to think that $P\&{\sim}Q$ cannot be 1-impossible despite being conceivable, this reason will count as a reason to think that the corresponding phenomenal and microphysical concepts cannot corefer despite being distinct epistemically rigid concepts after all.

No doubt, the above line of thought is well taken. It is, of course, true that if there is a good reason to think that $P\&{\sim}Q$ cannot be 1-impossible despite being conceivable, this will give us a good reason to think that phenomenal and microphysical concepts cannot corefer despite being distinct epistemically rigid concepts. But why think that $P\&{\sim}Q$ cannot be 1-impossible despite being conceivable? Whether $P\&{\sim}Q$ can be 1-impossible despite being conceivable comes down to whether the corresponding phenomenal and microphysical concepts can corefer despite being distinct epistemically rigid concepts. If this is so, then if we can show that there is no good reason to think that the corresponding phenomenal and microphysical concepts cannot corefer despite being distinct epistemically rigid concepts, we will have shown that there is no good reason to think that $P\&{\sim}Q$ cannot be 1-impossible despite being conceivable. But that is precisely what we have shown here. In effect the objection considered here does not work. Our response to the two-dimensional argument remains unthreatened.

I conclude again that premise 2 of the two-dimensional argument is unjustified. For this reason, the two-dimensional argument fails. Chalmers justifies premise 2 by the principle that ideal conceivability entails 1-possibility (CP), but given that premise 2 is unjustified, CP is unjustified as well. Speaking more generally, then, the two-dimensional argument fails because the principle CP is unjustified.

Acknowledgements

Thanks to an anonymous reviewer for very helpful comments.

Funding

I acknowledge the financial support of the National Science Centre, Poland/ project nr UMO-2017/25/B/HS1/02986.

References

Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Chalmers, David. 2002a. "Consciousness and Its Place in Nature." In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David Chalmers, 147–272. Oxford: Oxford University Press.

Chalmers, David. 2002b. "Does Conceivability Entail Possibility?" In *Conceivability and Possibility*, edited by Tamar Gendler, and John Hawthorne, 145–200. Oxford: Oxford University Press.

Chalmers, David. 2003. "The Content and Epistemology of Phenomenal Belief." In *Consciousness: New Philosophical Perspectives*, edited by Quentin Smith, and Aleksandar Jokic, 220–72. Oxford: Oxford University Press.

Chalmers, David. 2006. "Two-Dimensional Semantics." In *The Oxford Handbook of the Philosophy of Language*, edited by Ernest Lepore, and Barry Smith, 574–606. Oxford: Oxford University Press.

Chalmers, David. 2010. "The Two-Dimensional Argument against Materialism." In *The Character of Consciousness*, edited by David Chalmers, 141–92. Oxford: Oxford University Press.

Chalmers, David, and Frank Jackson. 2001. "Conceptual Analysis and Reductive Explanation." *Philosophical Review* 110 (3): 315–61. https://doi.org/10.1215/003 18108-110-3-315

Goff, Philip. 2017. *Consciousness and Fundamental Reality*. New York: Oxford University Press. https://doi.org/10.1093/oso/9780190677015.001.0001

Kallestrup, Jesper. 2006. "Physicalism, Conceivability, and Strong Necessities." *Synthese* 151 (2): 273–95. https://doi.org/10.1007/s11229-004-7325-9

Kaplan, David. 1989. "Demonstratives". In *Themes from Kaplan*, edited by Joseph Almog, John Perry, and Howard Wettstein, 481–563. New York: Oxford University Press.

Loar, Brian. 1997. "Phenomenal States". In *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan, and Güven Güzeldere, 597–616. Cambridge, MA: MIT Press.

Loar, Brian. 1999. "On David Chalmers's The Conscious Mind." *Philosophy and Phenomenological Research* 59 (2): 465–72. https://doi.org/10.2307/2653684

Papineau, David. 2002. *Thinking about Consciousness*. Oxford: Oxford University Press.

Papineau, David. 2007. "Phenomenal and Perceptual Concepts." In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin Alter, and Sven Walter, 111–44. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195171655.003.0007

Organon F 27 (3) 2020: 411–415
https://doi.org/10.31577/orgf.2020.27307
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

DISCUSSION NOTE

Limiting Cases of Modal Modification: Reply to Kosterec

Bjørn Jespersen*

Kosterec (2019) points out that my current theory of modal modifiers cannot deal satisfactorily with limiting cases. This note solves the problem. The form of the solution is to leave the existing theory as is and instead add a clause handling the limiting case which Kosterec brings up and another clause handling the limiting case at the other end of the spectrum.

My theory of modal modifiers, as set out in (2013), works well, as long as the argument property being modified is either (*i*) a purely contingent property or (*ii*) a contingent property with an essential core, provided the resulting modified property (*MF*) is not applied to an element of the essential core of *F*.[1] To stick with the original example of mine that Kosterec takes over, we treat this predication as a datum:

"Individual *a* is an alleged assassin"

Its analysis in Transparent Intensional Logic is this:

$$\lambda w \lambda t\, [[^0 Alleged\ ^0 Assassin]_{wt}\ ^0 a]$$

Types: $Alleged/((o\iota)_{\tau\omega}\ (o\iota)_{\tau\omega})$; $Assassin/(o\iota)_{\tau\omega}$; a/ι; $w/*_1\!\rightarrow_v \omega$; $t/*_1\!\rightarrow_v \tau$.

[1] See (Duží et al. 2010, §1.4.2.1) for the definitions of *purely contingent property* and *contingent property with an essential core*. See (*ibid.*) for notions and notation.

* Utrecht University and VŠB-Technical University of Ostrava
Department of Philosophy and Religious Studies, Utrecht University, Janskerkhof 13, 3512 BL Utrecht, The Netherlands; Department of Computer Science, VŠB-Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava, Czech Republic
b.t.f.jespersen@uu.nl

I claim that two conclusions are forthcoming. The first conclusion is that there is some property f which a is alleged to have:

$$\lambda w \lambda t\, [{}^0\exists \lambda f\, [[{}^0 Alleged\ f]_{wt}\ {}^0 a]]$$

Types: $f/*_1{\rightarrow_v}\ (o\iota)_{\tau\omega}$; $\exists/(o(o(o\iota)_{\tau\omega}))$.

This predication is non-trivial, because not all of us are being alleged to have some property or other.[2] The second conclusion is that maybe a is an assassin and maybe a is not an assassin:

$$\lambda w \lambda t\, [[{}^0 Alleged\ {}^0 Assassin]_{wt}\ {}^0 a] \supset$$
$$[{}^0\exists \lambda w'\, [{}^0\exists \lambda t'\, [{}^0 Assassin_{w't'}\ {}^0 a]] \wedge {}^0\exists \lambda w''\, [{}^0\exists \lambda t''\, \neg[{}^0 Assassin_{w''t''}\ {}^0 a]]]$$

where $w' \neq w''$, $t' \neq t''$.

A dichotomy is induced over the domain of world/time pairs, such that in one half of the domain it is true that a is one of the assassins and in the other half it is false that a is among the assassins. The open question is which side of the fence a given world/time pair of evaluation comes down on. The logical behaviour that the modal modifier displays is that it oscillates, as it were, between being *subsective* and being *privative*. A subsective modifier has the effect that the modifier is eliminated and the original argument property is predicated of the individual in question. For instance, a skilful surgeon is a surgeon. A privative modifier has the effect that the predication of the privatively modified property is replaced by the boolean negation of the predication of the argument property.[3] For instance, it is not the case that a fake banknote is a banknote.

The counterexample Kosterec levels against my theory is this predicate:

‘is an alleged discoverer of the highest prime number’

There is no highest prime number, hence nobody can instantiate the property of discovering the highest prime number, hence the left-hand conjunct

[2] See (Jespersen 2016) for the *general rule of left subsectivity*, which in (Duží et al. 2010, §4.4) was introduced under the name of *pseudo-detachment*.

[3] – in the case of *single* privation, that is. In the case of *iterated* privation, privative modifiers are replaced by the general privative modifier *Non*. See (Jespersen et al. 2017).

(from being an alleged assassin to being an assassin) is false, hence the conjunction is false, hence the inference is invalid.

When confronted with impossibilities, the strategy pursued by Transparent Intensional Logic is not to usher in impossible worlds as additional points of evaluation. Instead we introduce constructions of conditions that could not possibly be satisfied (see Duží et al. 2020). What we need here is, first of all, a construction of the impossible property of being a discoverer of the highest prime:

$$\lambda w \lambda t\, [\lambda x\, [{}^0\mathit{Discover}_{wt}\, x\, {}^0[{}^0\imath \lambda y\, [[{}^0\mathit{Prime}\, y] \wedge {}^0\forall \lambda z\, [[{}^0\mathit{Prime}\, z] \supset [{}^0\geq\, y\, z]]]]]]$$

Types: $x/*_1\rightarrow_v \iota$; y, $z/*_1\rightarrow_v \tau$; $\mathit{Discover}/(o\iota *_n)_{\tau\omega}$; $\mathit{Prime}/(o\tau)$; $\imath/(\tau(o\tau))$; $\forall/(o(o\tau))$; $\geq/(o\tau\tau)$.

The analysis of "a is an alleged discoverer of the highest prime" is:

$$\lambda w \lambda t\, [[{}^0\mathit{Alleged}\, \lambda w \lambda t\, [\lambda x\, [{}^0\mathit{Discover}_{wt}\, x\, {}^0[{}^0\imath \lambda x\, [[{}^0\mathit{Prime}\, y] \wedge {}^0\forall \lambda z\, [[{}^0\mathit{Prime}\, z] \supset [{}^0\geq\, y\, z]]]]]]]_{wt}\, {}^0a]$$

How do we eliminate *Alleged*? By invoking the fact that at no world/time pair is a, or anyone else, someone with the property of discovering the highest prime.

We are going to define the property X, which is an analytic property of ι-properties, namely the property of being necessarily uninstantiated ('empty'). Thus, its functional arguments being $F_i \in X$, *Alleged* modifies impossible empirical conditions.[4] First of all, we define $\varnothing^\iota$ as the set of empty ι-sets, whose respective characteristic functions do not return the truth-value 1 for any argument, i.e., they either return 0 or are undefined:

$${}^0\varnothing^\iota =_{df} \lambda e\, [{}^0\forall \lambda x\, \neg[{}^0\mathit{True}*\, {}^0[e\, x]]]$$

Types: $e/*_1\rightarrow (o\iota)$; $\varnothing^\iota/(o(o\iota))$; $=/(o(o(o\iota))(o(o\iota)))$; $\mathit{True}*/(o*_n)$: the set of such constructions as v-construct 1 for every valuation v.

[4] I should stress that the addition to the theory of modal modifiers I have offered here still does not extend to purely arithmetical cases as expressed by predicates like 'is an alleged proof of the continuum hypothesis'. What is already clear, though, is that, *Proof* being of type $(o*_n)$, namely, a set of hyperpropositions, *Alleged*' as denoted in 'is an alleged proof' must be of type $((o*_n)\,(o*_n))$.

Now define X as follows:

$$^0X ='_{df} \lambda f\,[^0\forall\lambda w\,[^0\forall\lambda t\,[^0\varnothing^{\iota}\,f_{wt}]]]$$

Types: $X/(o(o\iota)_{\tau\omega})$; $f/*_1\rightarrow_v (o\iota)_{\tau\omega}$; $='/(o(o(o\iota)_{\tau\omega})(o(o\iota)_{\tau\omega}))$.

A parallel definition of X, to be deployed below, is this one:

$$^0X ='_{df} \lambda f\,[^0\forall\lambda w\,[^0\forall\lambda t\,[^0\forall\lambda x\,\neg[^0True_{wt}\,\lambda w\lambda t\,[f_{wt}\,x]]]]]$$

Type: $True/(oo_{\tau\omega})_{\tau\omega}$: the empirical property of truth-conditions/$o_{\tau\omega}$ of being satisfied (i.e., returning 1) at a given world/time pair of evaluation.

Where $F\in X$, the elimination of *Alleged* proceeds as follows:

$$^0\forall\lambda w\,[^0\forall\lambda t\,[^0\forall\lambda x\,[[[^0Alleged\,^0F]_{wt}\,x]\supset\neg[^0True_{wt}\,\lambda w\lambda t\,[^0F_{wt}\,x]]]]]$$

This clause is the solution to the problem presented by the first limiting case, which Kosterec has brought up. Of course, X has a mirror-image, Y, which is the analytic property of ι-properties of being necessarily instantiated ('trivial'):

$$^0Y ='_{df} \lambda f\,[^0\forall\lambda w\,[^0\forall\lambda t\,[^0\forall\lambda x\,[f_{wt}\,x]]]]$$

The elimination of *Alleged* now proceeds as follows, for any $G\in Y$:

$$^0\forall\lambda w\,[^0\forall\lambda t\,[^0\forall\lambda x\,[[^0Alleged\,^0G]_{wt}\,x\supset[^0G_{wt}\,x]]]]$$

The addition of the first clause to my theory of modal modification departs from the observation that nobody and nothing could possibly instantiate any property F when $F\in X$ and F has been modified by *Alleged*. Remember that the definition of modal modifiers applicable to contingent empirical properties (i.e., (*i*), (*ii*)) embodies a bifurcation, and that the way the cookie happens to crumble determines whether the alleged property is true of the individual in question. Modal modifiers applicable to necessarily uninstantiated properties are importantly different, in that the predication can go only one way: the alleged property must fail to be true of the individual in question. Therefore, my account of this second category of modal modifiers aligns them formally with privative modifiers. The difference between the two, though, is that the source of privation is not the modifier (*Alleged* versus *Fake*), but the argument property itself (*being a discoverer of the highest prime*, *being a married bachelor* versus *being a banknote*). At the other end of the spectrum, when $G\in Y$ and G has been modified by

Alleged, everyone and everything must instantiate G, thus aligning this third category of modal modifiers with (trivial) subsective modifiers. The difference between the two is that the source of triviality is, likewise, not the modifier (though *Genuine*, as in *being a genuine diamond*, adds or detracts nothing), but the argument property itself (e.g., *being as tall as one is*).

Acknowledgments

I am indebted to Marie Duží for valuable comments.

Funding

This research was funded by the Grant Agency of the Czech Republic (GACR) project GA18-23891S *Hyperintensional Reasoning over Natural Language Texts*.

References

Duží, Marie, Glavaničová, Daniela, and Jespersen, Bjørn. 2020. "Impossible Individuals as Necessarily Empty Individual Concepts." In *Logic in High Definition: Developing Fine-Grained Semantics*, *Trends in Logic*, edited by Alessandro Giordani and Jacek Malinowski, forthcoming.

Duží, Marie, Jespersen, Bjørn, and Materna, Pavel. 2010. *Procedural Semantics for Hyperintensional Logic. Foundations and Applications of Transparent Intensional Logic,* LEUS, vol. 17, Heidelberg et al.: Springer-Verlag. https://doi.org/10.1007/978-90-481-8812-3

Jespersen, Bjørn. 2016. "Left Subsectivity: How to Infer That a Round Peg is Round." *dialectica* 70 (4): 531–47. https://doi.org/10.1111/1746-8361.12159

Jespersen, Bjørn, Carrara, Massimiliano, and Duží, Marie. 2017. "Iterated Privation and Positive Predication." *Journal of Applied Logic* 25 (suppl.): 48–71. https://doi.org/10.1016/j.jal.2017.12.004

Jespersen, Bjørn and Primiero, Giuseppe. 2013. "Alleged Assassins: Realist and Constructivist Semantics for Modal Modifiers." In *Lecture Notes in Computer Science*, vol. 7758, edited by Guram Bezhanishvili, Sebastian Löbner, Vincenzo Marra, and Frank Richter, 94–114. https://doi.org/10.1007/978-3-642-36976-6_8

Kosterec, Miloš. 2019. "Contradiction of Modal Modification. *Organon F* 26 (2): 298–300. https://doi.org/10.31577/orgf.2019.26207

Organon F 27 (3) 2020: 416–420
https://doi.org/10.31577/orgf.2020.27308
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

BOOK REVIEW

Neil Roughley and Kurt Bayerts (eds.): *The Normative Animal?* New York: Oxford University Press, 2019, x+380 pages

Jaroslav Peregrin*

Let me start this review in a personal tone. In a paper I wrote some ten years ago ('Rules as the Impetus of Cultural Evolution,' *Topoi* 33, 2014, 531–45), I ventured the following thought:

> There are many suggestions about what makes us humans special: *soul*, *mind*, *language*, *culture*, *reason* .... In this paper I have indicated that we may characterize man as a normative being. Not that this proposal by itself would be original—of course it goes back at least to Kant; and recently a persuasive case for it has, in effect, been made by Brandom. However, I have tried to show that if we accept the analyses of the concept of *rule* put forward by Sellars, we can embed this characterization in the evolutionary stories of how we humans have become what we are.

It was at that time that it came to me that the characterization of us humans as *normative beings* cuts surprisingly deep, firstly because many other of our specific abilities somehow depend on our normativity and, secondly, because this normativity is potentially explainable in a naturalistic manner. Hence my great excitement when I found that Oxford University Press have now published a volume called *The Normative Animal?*, as I assumed it would be devoted to this very idea. And indeed, one of the volume's editors, Neil Roughley, starts his introductory text as follows:

* Institute of Philosophy of the Czech Academy of Sciences & Faculty of Philosophy, University of Hradec Králové
Institute of Philosophy of the Czech Academy of Sciences, Jilska 1, 110 00 Prague 1, Czech Republic; Faculty of Philosophy, University of Hradec Králové, Rokitanského 62, 500 03 Hradec Králové, Czech Republic
peregrin@flu.cas.cz

> Humans, it has often been claimed, are characteristically or even essentially rational, linguistic, social, or moral creatures. If these characterisations are intended to name the essence or nature of being human, however understood, then they would appear to be in conflict. This volume is built around the question of whether these characterisations may not turn out to be compatible because they all ground in a more basic feature: that of being creatures whose lives are structured at a fundamental level by their relationships to norms. The various capacities singled out by talk of rational, linguistic, social, or moral animals might then all essentially involve the orientation to obligations, permissions, and prohibitions. If this is so, then perhaps it is a basic susceptibility, or proclivity to the normative or deontic regulation of thought and behaviour that enables humans to develop the various traditionally emphasised features of their life form.

This, obviously, echoes my own musings about the import of normativity, so it fueled my keenness to see whether the papers assembled in the volume manage to put such musings on a more solid foundation. After reading the book, my impression is that they do in some partial areas, but that the question about the depth of the fundamental level (as referred to above in the introductory text) is not broached quite satisfactorily yet.

In the introductory essay which, together with a general introduction to the papers collected in the volume, constitutes the first part of the book, Neil Roughley provides a useful disambiguation of the concept of "normative creatures," namely as

(NA1) The creatures are beings to whom the norms *apply*.
(NA2) The creatures *regulate* their behaviour in line with what they take to be normatively required of them.
(NA3) The creatures are in some sense *creators* and *upholders* of the norms.
(NA4) The creatures are *enforcers* of the norms.

He himself considers (NA2) as crucial. (Here, I must admit, I disagree: personally, I would go for (NA3), which, in my view, involves (NA4). If one opts for either (NA1) or (NA2) as crucial, then a separate explanation is needed to account for how the norms we humans obey ever came into being.) Roughley also discusses the three kinds of norms to which the ensuing book parts are devoted: social norms, moral norms, and linguistic norms.

Of the papers in the next part, devoted to social norms, the last is the one I find the most interesting. It is written by two prominent members of the former Leipzig group around Michael Tomasello, namely Marco Schmidt and Hannes Rakoczy, who have been studying the ontogenesis of normativity for over a decade. In their current contribution ('On the Uniqueness of Human Normative Attitudes'), they summarize the results of their experiments with children and draw some important conclusions. The concept they focus on is that of having a *normative attitude*: the tendency to divert fellow beings either away or toward various courses of action. It is the awakening of these attitudes that the authors claim marks the early ontogenesis of us humans and makes us the normative beings we are. I think this empirical approach to normativity has a dual significance: on the one hand, normative attitudes can be seen as constitutive of implicit rules which, in turn, can be seen as underlying all rules; while also, on the other hand, they can be seen as something explainable in purely naturalistic terms.

Another interesting paper in this part is the first contribution by Peter M. Kappeler, Claudia Fichtel, and Carel P. van Schaik ('There Ought to Be Roots: Evolutionary Precursors of Social Norms and Conventions in Non-Human Primates'). These authors, too, stress the importance of normative attitudes (though they do not call them so): they suggest that normativity consists in the expectations of individuals in respect to the courses of actions of others, and claim that such expectations can be diagnosed only in terms of the perceivable reactions of the individuals to the violations of such expectations. The authors also try to identify the pre-normative components out of which our normativity has been assembled; which they see especially in dyadic interactions.

This part of the book also gives us contributions from two other authors, Christoph Antweiler ('On the Human Addiction to Norms') and Karl Mertens ('On the Identification and Analysis of Social Norms and the Heuristic Relevance of Deviant Behaviour'), who address normativity and its role within human communities mostly from the viewpoint of philosophy and cultural anthropology.

The third part of the book, devoted to moral norms, cannot fail to reflect the ongoing discussion concerning the question whether moral norms constitute a specific kind of norm or whether they are merely a loosely delimited "core" of social norms. That moral norms are *sui generis* is defended by this part's introductory paper, 'The Evolution of Human Normativity: The Role of Prosociality and Reputation Management' by Carel P. van Schaik and Judith M. Burkart.

The authors argue that moral norms are the only *genuine* norms, the existence of which was necessitated by the organization of the communities of our ancestors, all other norms then arising as their relaxed variants. Karl Bayertz, in the next contribution ('The emergence of moral normativity'), defends the opposite standpoint: according to him, moral norms have developed out of the more broadly social ones especially in connection with the onset of language. Something similar is claimed also by the author of the following paper ('Joint Activities and Moral Obligation'), Holmer Steinfath, who, likewise, stresses the role of language within the emergence of moral norms out of the social ones. He is convinced that the importance of language consists especially in that it allows us to produce justifications. The last two chapters of this part of the book ('The Development of Domains of Moral and Conventional Norms, Coordination in Decision-Making, and the Implications of Social Opposition' by Elliot Turiel and Audun Dahl, and 'Moral Obligation from the Outside In' by Neil Roughley), argue that social and moral rules are quite different species, neither of them being an outgrowth of the other.

The fourth and last part of the volume, devoted to linguistic rules, starts with a contribution by Nicola Kompa ('Language evolution and linguistic norms'), who argues that it is rules that make it possible to turn signals into symbols (to allow them to acquire a conventional meaning that is stable) and in this sense they underlie language. A similar picture is drawn by N. J. Enfield and Jack Sidnell ('The normative nature of language'), who also use the Peircean conceptual framework. They claim that "whenever a word is used, that word will dependably invoke a definable core idea in the minds of people who hear the word being used" and insist that "the norm-governed flow of sequences of social interaction provides a matrix in which norms of language are used for regimenting the use of language, and thereby concretizing these norms in the form of semantic conventions." Anne Reboul ('Can there be linguistic norms?'), in contrast to this, rejects normativity as a substantial ingredient of language. In the last paper of this part of the book, Hanjo Glock's 'The normativity of meaning revisited,' the author reviews and elaborates on some philosophical arguments, going back to Wittgenstein, for the normativity of meaning.

The question whether we humans are normative animals can be construed in various, weaker and stronger, senses. In the weakest sense, the question asks whether we are able, perhaps in contrast to other animals, to abide by norms; and needless to say that to this the answer is positive. In a stronger sense, the question asks whether normativity is a ubiquitous feature of human life form.

Here the book brings forth a lot of evidence for the positive answer to this question, while also presenting us with some dissenting voices, which force the "normativists" to refine their positions. Importantly, though, I would hold that this question can be asked in a yet stronger sense, asking whether rules are our *principal* distinction, underlying all other specifica of us humans, such as reason, language, cognition, morality etc. Personally, I believe that even here the answer may be positive, and I find it fascinating to probe this topic. (It is not just a lunatic idea: inferentialism has come up with an elaborated theory regarding language and linguistic meaning, according to which meaning is a matter utterly of rules, and hence that language is underlain by rules; and views that our specifically human cognition is derived from language are no longer counted as extreme.) In this respect I think the book does not go far enough (though, we saw, one of the editors in his introduction, does interpret the question in this strong sense), perhaps because most of the authors do not think that the strong sense of the question is worth being taken seriously.