Croatian Journal of Philosophy

Vol. XVIII · No. 53 · 2018

Croatian Journal of Philosophy

1333-1108 (Print)
1847-6139 (Online)

Co-published by
"Kruzak d.o.o."
Naserov trg 6, 10020 Zagreb, Croatia
fax: + 385 1 65 90 416, e-mail: kruzak@kruzak.hr
www.kruzak.hr

and

Institute of Philosophy
Ulica grada Vukovara 54/IV, 10000 Zagreb, Croatia
fax: + 385 1 61 50 338, e-mail: filozof@ifzg.hr
www.ifzg.hr

Available online at http://www.ceeol.com and www.pdcnet.org

CROATIAN
JOURNAL
OF PHILOSOPHY

Vol. XVIII · No. 53 · 2018

Articles

Book Reviews

Introduction

The first five papers of this issue of the Croatian Journal of Philosophy *presents the proceedings of a mini conference with Michael Devitt that was held in Rijeka in April 2017 under the title 'Linguistic Intuitions and Natural Kinds' organized by the Croatian Society for Analytic Philosophy. The other four papers are mainly concerned with different aspects of rationality.*

Nenad Miščević's *paper "Intuitions: Epistemology and Metaphysics of Language" deals with epistemology of linguistic intuitions, and defends a moderate Voice-of-competence view in discussion with Michael Devitt, who sees them as products of general intelligence or Central Processing Unit. The second part of the paper deals with validity of linguistic intuitions and offers a compromise solution: linguistic intuitions are valid because their object, the standard linguistic entities, are production -and response-dependent. The solution is briefly situated on the map of general response-dependentism.*

Michael Devitt's *and* Dunja Jutronić's *papers are direct responses to Nenad Miščević's articles "Reply to Michael Devitt", and "Reply to Dunja Jutronić", both published in 2014. Miščević defends a modified version ("MoVoC") of the received view that these intuitions are the product of a linguistic competence.* Michael Devitt *has always rejected all versions of the received view urging instead that intuitions are, like perceptual judgments, empirical theory-laden central-processor responses to phenomena. He emphasizes here, against Miščević, that the claim about a speaker's intuitions about strings is not to be conflated with a claim about her understanding of strings. Furthermore, he develops his claim that Miščević's MoVoC is implausible in three important aspects. Devitt finally points out that these are not the main problems for MoVoC.*

In her contribution "Intuitions Once Again!" Dunja Jutronić *first presents some of her most important answers to Miščević's objections to her 2014 paper. Secondly and more importantly, she points out that there is a possible confusion or misunderstanding about the distinction between the object-level (sentence produced) and meta-level (sentence judged). Jutronić argues that competentionalist actually conflates object and meta levels and shows the final consequences of such a conflation Finally, she briefly comments on the so-called 'Route Question', that is the path from the underlying competence to the central processor and argue that Miščević, or any competentionalist, cannot provide an explanation for it.*

Zdenka Brzović's *paper "Devitt's Promiscuous Essentialism" examines Devitt's version of essentialism, a view that stirred a lot of debate amongst philosophers of biology by going against the mainstream view of "death of essentialism" in evolutionary biology. Brzović goes through the main tenets of the essentialist view, examines the relation between Devitt's view and the so-called traditional essentialism, and the cluster approaches to natural kinds. She concludes that Devitt holds a very flexible variety of pluralistic essentialism that she terms promiscuous essentialism.*

Urška Martinc *in her paper "Devitt's 'Intrinsic Biological Essentialism'" primarily focuses on Michael Devitt's article "Resurrecting Biological Essentialism". Using examples from biology and the analogy from some examples from chemistry, Martinc analyses advantages and disadvantages of Devitt's arguments for intrinsic biological essentialism.*

In "Structured Propositions, Unity, and the Sense-Nonsense Distinction" Octavian Ion *starting point of discussion is Herman Cappelen (2013) who thought that we need to rethink and reintroduce the important distinction between sense and nonsense that was neglected during Logical Positivism's demise. However, Ion's delineation of the bounds of sense is different from Cappelen's. One of his main goals is to argue that category mistakes are paradigmatic examples of nonsensical sentences.*

Martina Blečić *in her contribution "Do Conversational Implicatures Express Arguments?" suggests that the idea that conversational implicatures express argument can be significant for the notion of communicational responsibility. It is proposed that we should consider conversational implicatures as reason-giving arguments in which the speaker (arguer) addresses a hearer who does not need to reply. In such cases, the speaker is not trying to convince the hearer to accept his position but is explicitly stating a reason in support of his intended message. Blečić argues that her approach can strengthen the idea of the speaker's communicational responsibility for an implicated message even in the case when he wants to distance himself from it.*

Conspiracy theories seem to play an increasing role in public political discourse, and Daniel Cohnitz's *paper "Conspiracy Paranoia: On the Rationality of Conspiracy Theories" urges that we should find out why conspirational thought is recently gaining such support and how to respond to it. People who believe in conspiracy theories are often ridiculed as nutcases or paranoid crackpots, while they portray themselves as particularly critical, better informed and enlightened responsible citizens. One of the central questions that needs to be answered here is what mindset leads to the believe in conspiracy theories? Finding out which of the two above mentioned characterizations is correct, is crucial for coming up with the appropriate response to the rise of conspirational thought.*

In his paper "Wisdom and Reason" Andrei Mărăşoiu presents Ryan's (2012) theory of wisdom as deep rationality. Namely, to believe or act

wisely is to believe or act in a justified way, informed by a body of other justified beliefs about the good life. Ryan elaborates the view along evidentialist lines: one's belief or act is justified when it is based on the best available evidence. Mărăşoiu points to a number of counterexamples to this approach and he argues that, instead of evidentialism, Ryan's view should include virtue theory, which helps explain the seeming counterexamples. He focuses on the virtues of openness to experience, and of steadfastness in the face of experience.

DUNJA JUTRONIĆ

Croatian Journal of Philosophy
Vol. XVIII, No. 53, 2018

Intuitions: Epistemology and Metaphysics of Language

NENAD MIŠČEVIĆ
University of Maribor, Maribor, Slovenia

The paper addresses the issues about grammatical intuitions in a programmatic sketch. The first part deals with epistemology of such intuitions and defends a moderate Voice-of-competence view in discussion with Michael Devitt, the ordinarist, who sees them as products of general intelligence or Central Processing Unit. The second part deals with the problem for their validity and offers a compromise solution: linguistic intuitions are valid because their object the standard linguistic entities, are production -and response-dependent. Competence does dictate what is correct, and what is not, the order of determination goes from the internal to the external, or external-seeming language items. An external token string has linguistic properties because it would be interpreted as having them by the normal language-hearer and would be produced by a process that would form it respecting the nature of these properties. The solution is briefly situated on the map of general response-dependentism.

Keywords: Intuition, competence, ordinarism, response-dependence.

1. *Introduction*

The paper is dedicated to Michael Devitt (to be called just "Michael" in the sequel), and continues our long and fruitful discussion about linguistic, in particular syntactic, intuitions. It has two parts, the first more epistemological, but on the non-normative side, the second more metaphysical. Let me say a few words about each. The main epistemological debate in philosophy of language concerns people's linguistic intuitions. Let me borrow an example from Isaac (2008: 178) to illustrate the kind of items that will be discussed in the paper. Suppose a linguist confronts John, a native speaker, with a following sentence:

(H*) "Herself loves Mary".

Is this a sentence of your language, the linguist asks. No, answers John. John has clear feeling that something is amiss with (H*), and the feeling gives rise to the judgment and belief expressed by his report "No". The feeling, the belief-state and the judgment, and sometimes even the report (all of them, or at least some of them) are called "intuition"; I will reserve the term for the first three of them. A famous recent tradition in philosophy of language and linguistics, initiated by Chomsky takes intuitions as the main source of data for the linguist. The tradition equally sees them as products of linguistic competence; the term is ambiguous between the ability and the mechanism, so I will use it for both, but most often for the later. I agree with the general idea, but I would like to allow that many opinions that people voice as their "intuitions" contain a lot of material not produced by, and therefore not revelatory of, the pure linguistic competence. (I have been defending similar views about other competences like the, logical and the spatial-geometrical ones). I have been calling my own line "the Moderate Voice-of-competence view", since it takes competence as basic, but allows for very strong interferences or external contributions to the production of intuition. John's intuition at best points to a discreet voice of competence. So, the Moderate Voice-of-Competence view that I would like to develop and briefly defend here, claims that intuitions form a kind, albeit relatively superficial one, shearing their phenomenal properties classically described by Descartes in his *Regulae* and *Principles*. in terms of "clear and distinct cognition", of being „present and apparent to an attentive mind, (Principle XLV, 168). And there is a capacity, or rather several of them associated with intuition. Further, they are extroverted, turned towards the items they are explicitly about, and normatively answerable to them, since they want them to teach us about things "outside" (possibly in Platonic heaven), not merely about our representation(s) of them). The view also takes seriously the actual dialectics of having intuitions: asking (or being asked) a question, then going through imagining a scenario, if necessary putting oneself in the shoes of the imagined person (or a person-like entity, say, zombie), and then giving a simple, preliminary answer to the question, formulating the immediate intuition, often to be developed by considering other examples, and so on. This intuition-related effort involves a lot more than mere inference following rules of logic. Further, the view is for the most part committed to realism about the objects of intuitions, and in the second part of the paper I develop a moderately realistic view about linguistic entities. The view is very keen on the explainability of intuitions. Finally, it offers a somewhat complex answer about their normative epistemic status, tilted towards aposteriority: although intuitions are prima face a priori, their reflective justification has a rich structure in which a posteriori elements play a crucial role.

The theories on the opposite end favors deny the specific nature of intuitions, insist on holistic non-a priori justification, and see them as

products of general intelligence. Let me call this pole "ordinarism"; it has been invented long time ago in ethics, with Ewing (1971) and other moral intuitionists, like More (1991), in some of their moods), but is being now re-discovered for other domains, generalized and put forward in vigorous manner by Michael (Devitt 2005, 2006, 2012), and specifically for philosophical intuitions T. Williamson (2007).

I disagree with Michael's mistrust. He is an empiricist about linguistic intuitions, and he is comparing them to expert judgments, paleontologist in the field searching for fossils. She sees a bit of white stone sticking through grey rock, and responds immediately "a pig's jawbone." (Devitt 2012: 560), art experts correctly judging an allegedly sixth-century Greek marble statue to be a fake; of the tennis coach, Vic Braden, correctly judging a serve to be a fault before the ball hits the ground (Devitt 2006a: 104).

The second part of the paper deals with the further problem: where does the validity our intuitions come from? In order to answer it, it briefly visits the location problem for language, whether it is essentially E-language situated in the outside world, or I-language, inhabiting only the mind of speaker-hearer) and offers a compromise solution: standard linguistic entities are production- and response-dependent. An external token string has linguistic properties because it would be interpreted as having them by the normal language-hearer, and would be produced by a process that would form it respecting the nature of these properties. The solution is then briefly situated on the map of general response-dependentism.

2. *Linguistic intuitions: The voice of competence*

2.1. *Intuitionism-competentinalism*

Let us then start from scratch. The typical context in which linguists speak about intuitions is the one of linguistic research. The linguist presents a string of sounds (phonemes, letters) to the native speaker of the language investigated (often to oneself, if the language is one's mother tongue), and asks her to decide if this would be a sentence of her language. The immediate judgments prompted by the question are described as linguistic intuitions.

I shall here re-use my older examples, since I have changed my explanation at one important point, thanks to Michael, and I want to stress the continuity and discontinuity with my former presentations. To have a handy example I have borrowed from a fine introduction to Chomsky by John Collins (2008) a pair of sentences testing the predicate nominal agreement:

(W) They want to be teachers.
(W*) *They want to be teacher.

Imagine a native speaker, Ann, accepting the first and rejecting the second. Perhaps Ann sort of rehearsed the sentences in her inner fore,

asking herself whether she would say them, simulating actual saying, as the standard description goes. The next example is our already mentioned

(H*) "Herself loves Mary".

rejected in our story by the naïve subject John. Finally, here is the co-reference example.

(M) Mary knows that Jane loves herself.

(M) can be taken in two ways: the incorrect one, according to which Jane loves Mary, and the correct one, which is obvious. So, suppose the linguist asks the naïve subject John two questions and receives the following answers:

Q: Does Mary know that Jane loves her, Mary?
A: Of course, not.

Q: So, whom does Jane love?
A: She loves herself, Jane.

So, how are Ann's and John's cognitive apparatuses arriving to the verdict? Following the lead from the Chomskyan tradition, I would claim that it is mobilizing the particular competence, i.e. the same cognitive resource that produces or fails to produce similar sentences in real-life speaking. It is the competence itself that is doing the work, the central processor at best just passively reports the verdict of the competence, which is the intuition. Michael would claim that Ann's and John's apparatuses are mobilizing the cognitive resource that is normally in charge of understanding sentences. Ann's resource outputs the verdict Yes for the first sentence (W), and No, for the second (W*), in some neural code. John's outputs a No for the (H*) sentence. If the resource is competence (the particular, i.e. linguistic one, then the result is the voice of competence.

I will take is as agreed by all sides that the first stage must be a tentative production of the sentence, and I would add that it is being rehearsed and analyzed by Ann's competence. I mentioned that competence presumably comes out with some kind of answer, some Yes or No signal. I will argue that this is the most important element, the core, of the final intuition. The next stage is empirical theorizing at sub-personal level; Ann's central processor, CP for short, has to interpret the message, decides how to treat it, and then translate the message into the spontaneous belief, what we call intuition. The rest is reporting, producing the verbal output. This linguistic intuitional output has a very narrow range. Intuitions reported are formulated in an austere vocabulary, featuring mainly "acceptable" vs. "non-acceptable". So what does this tell us about folk-concept of grammaticality? Distinguish an egocentric minimal concept, expressed by "I wouldn't say S"/ "I would say S" (for some target sentence S), from rich socio-centric concept IS A PART OF MY COMMUNITY LANGUAGE, predicated of "S". Two related points: first, for producing relevant intuition, only the ego-

centric minimal concept is needed. Second, this concept is not clearly empirical, culled from past experience with one's own sayings. Its application is typically guided by immediate promptings of competence: if you ask me about a sentence in my mother tongue, Croatian, what I primarily do is that I either just "hear" that this is not what I would say, or I try to produce the sentence internally. My verdict is then an immediate Yes (or No, or, in the worst case Yes-and-No, e.g. if I am very drunk). And the basic data for the linguist are that Ann would say such-an-such and would mean such-and-such by the given expression. The rest is sociological theory. Remember, when a Chomskyan speaks about one's language, he means one's idiolect. Ann's opinion whether she is a typical representative of a community, and whether her English is good English are beside the point. Does Ann have to think that she is competent in the very language (idiolect) she is speaking? The very question sounds ludicrous. Why is this relevant? Because stating the judgment about a sentence in one's idiolect requires far less theory, if any at all, than reflection about social usage.

Let me rerun a simple scheme of the hypothetical production of a syntactic intuition I proposed a decade and half ago (Miščević 2016), with an important correction inspired by Michael's criticism. Here are two sentences

(W) They want to be teachers.
(W*) *They want to be teacher.

Imagine a native speaker, Ann, accepting the first and rejecting the second. Ann can rehearse the sentences in her inner fore, simulating producing it, or her linguistic cognitive apparatus can just analyze the sentence heard. Or, she can go on asking herself whether she would say them, simulating actual saying, as the standard description goes.

1	2	3	4
Processing the target sentence (heard or simulated)	*immediate, spontaneous verdict by specialized* competence *(intuition core)*	*empirical testing at sub-personal level*	*intuition*

In my earlier presentations of the MoVoC theory I have placed in the first box only simulation; Michael has kindly pointed out to my mistake, so I have enriched the content of the first box in in the meantime. The immediate spontaneous answer is the datum used by the central processor to arrive at the belief state, intuition proper. The generation of linguistic intuition-states seems to be rather isolated, independent of general intelligence, employed in stage 3 theorizing. of our flow-chart. "CP" stands for "central processor", the general intelligence.

Similarly, for more complicated structures, like the ones involving co-reference.

(M) Mary knows that Jane loves herself.

The immediate spontaneous answer is the datum used by the central processor to arrive at the belief state, intuition proper. The generation of linguistic intuition-states seems to be rather isolated, independent of general intelligence, employed in stage 3 theorizing. of our flow-chart. Second, it seems that the immediate, spontaneous answer (the item no. 2 on the flow chart) , is the item that deserves to be called "intuition". It is not necessary that verdict is unconscious; and if and when it is conscious, it seems to be the prime, if not the unique candidate for immediate (and perhaps obvious and compelling) judgment. In that case, the empirical testing proposal is just a verbal maneuver; calling the last stage, no. 4 "intuition", instead of stage no. 2.

On the view I propose, "the Moderate Voice-of-competence view", the answers in no. 2, no. 4. and indirectly in no. 5 are often produced by, and in such a case, revelatory of the linguistic competence. Only the competence has access to grammar, whatever its nature.

The ordinarist theories all this, since they want to deny the specific nature of intuitions. The ordinarist's flow-chart has roughly the following shape:

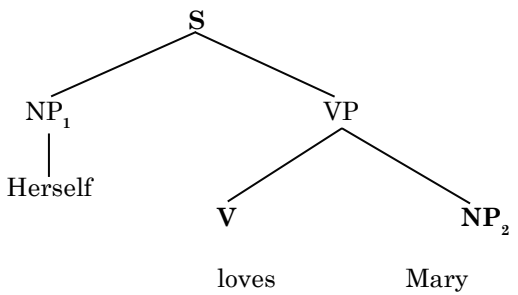
1	2	3
Processing the target sentence (heard or simulated)	*empirical testing* by central processor	*intuition (so-called)*

The crucial question is whether and how the central processor, i.e. general intelligence accessing memory, can quickly analyze complicated syntactic structures involved in ordinary sentences, like the ones involving co-reference. The empiricist-ordinarist answer is that the capacity is due to empirical exercise, like the one characterizing the expert knowledge of, say, a paleontologist, or a good tennis player. (I call it ordinarist, since it fits nicely with other views that see intuitions as items of ordinary knowledge, most famous of which is the view of Tim Williamson (2008). Linguistic ordinarism denies that seemingly linguistic judgments form a significant epistemic kind, intuitions, and that there is a distinct capacity producing them.

Roughly, the ordinarist hopes that the contribution of the competence is minimal, and the holistic contribution of CP maximal and essential. For her, intuitions are basically the products of holistic theorizing, not of special, dedicated competence. This is why they are not special, why they contain so much empirical material, and why it is wrong to take them to be a priori. I find it incredible. Nothing in our ordinary empirical knowledge points to a general structure-recognizing empirical ability of this power. And, in his forthcoming book Devitt (2019) is quite skeptical about people's semantic intuitions! He is not

rejecting them altogether, but does not want to use them as evidence, and criticizes others for using them exclusively as evidence. One can read him in a more radical and in a more moderate way: on the radical reading, he is rejecting them altogether, on the moderate reading he is just demanding the theorist to re-check her intuition appealing to non-intuitional sources, like elicited production or analysis of the corpus. In the discussion, Michael opted for the moderate reading, but some of his formulation suggest the radical version. I will say more about the variants of the ordinarist view in the sequel. In order to determine who is right, we have to discuss the proposals following the flow-chart(s) stage by stage.

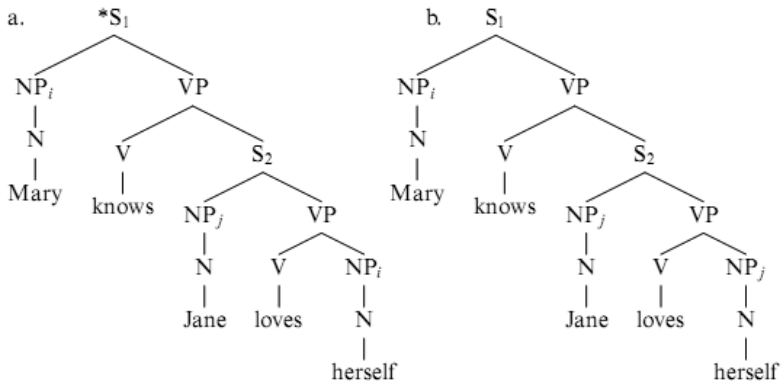
First, our examples of simple contrast between what is syntactically correct as oppose to incorrect. Take the sentence (H*) "Herself loves Mary", offered to John, and rejected by him. Why did he reject it? What is wrong with the sentence, that otherwise looks symmetrical to "Mary loves herself", which is, let us suppose, OK for John? The answer is very simple for the professional linguist (Isaac 2008: 178), but less so for John's general intelligence (and for mine as well, for that matter). The anaphor "Herself" has to be bound. In order for "Herself" to be bound, "Mary" should c-command it. However, the structure of the sentence, somewhat simplified, looks like this.



So, it is not the case that "Mary" c-commands "Herself". In order to produce these answers, John's internal parsing device should obtain the information that "Herself" is not syntactically connected to "Mary" in the right way. How can parser arrive at this? This can happen in two, or even three ways. First, something resembling the tree, call it "mental phrase marker" (I learned the term from David Pereplyotchik, so thanks go to him) can be either implemented in a non -explicit, non-representational way in the parser, so when the parser runs the parsing operation the result is that the sentence does not fit the mental phrase marker; the operation is either aborted, or some "red light" signal is emitted. Alternatively, the mental phrase marker is a full, explicit representation, and the parser "draws" it, in the way we did it here. Finally, and least probably, some items in the marker are merely implicit, others explicit.

Let us go along with the ordinarists and grant them the hypothesis that the mental phrase marker is implemented in an implicit way, so that there is no explicit representation in John's parser isomorphic to the tree we draw. If this is the case, what is the output available to the general intelligence?

We now pass to co-reference. Remember the sentence "(M) Mary knows that Jane loves herself". The linguist has put to the naïve subject John two questions, first, Does Mary know that Jane loves her, Mary, and the second "So, whom does Jane love?". The first question has been answered in the negative, and the second prompted the answer one would expect, namely that she loves herself, Jane. In order to produce these answers, John's internal parsing device should obtain the information that "herself" is in a right and complicated way connected to Jane, and not to Marry. It should have had at its disposition something corresponding to our two trees.



The crucial difference between them is the following: In the monster tree on the left-hand side, the first NP is co-indexed with the second NP containing the anaphor *herself*, which it also c-commands, and therefore binds, but the minimal clause containing the second NP is wrongly chosen: it is not the first, but the second one, S_2. So, whereas the conditions for binding are fulfilled, the locality condition is violated. (I am retelling the explanation from the textbook). The second tree fulfills both conditions: "Mary" binds "herself" (co-indexing and c-command are satisfied), and they occur together in the minimal clause, which is S_1 itself. So, everything is legal. How can parser arrive at this? Again, this can happen in two to three ways, non explicit, non-representational way fully representational or mixed. The analogous story can be told for "They want to be teacher".

2.2. *The intuition-core*

Suppose we all agree that competence involves at least embodied and non-represented rules, and operates according to them. We also agree

that the immediate production is a datum. Michael and I went through several forms which the contrast between his ordinarism and my competentinalism might take, concentrating upon stages 2 and 3 of each view. First, the immediate answer of the competence, stage 2, and its origin, the tentative production of the sentence at stage 1. What does the immediate answer consist in?

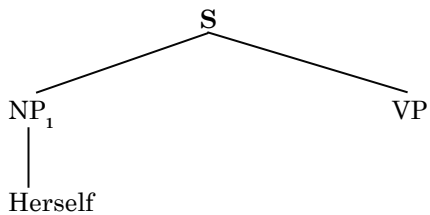
Dunja Jutronić mentioned once the extreme possibility that there is almost no answer at all, that competence just reiterates the sentence proposed, say "They want to be teacher." That won't do, since this is no new datum at all.

The second possibility has been put forward in Devitt's answer to my criticism, in the context of discussing the question whether person's answer to the linguist's question is the datum:

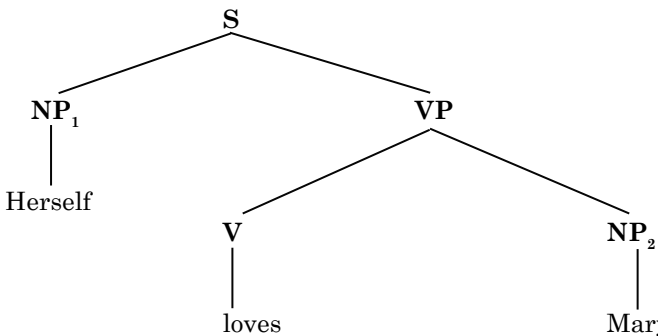
> Her answer is not /i.e. the datum-NM/; it is part of the central-processor reflection. The datum is the experience that the answer is about. (2006c: 594 fn.22)

The experience, as made clear by the context, is "the experience of simulating the behavior" (2006: 594, the body of the text), i.e. the neural-verbal behavior of producing or trying to produce the target string.

This is hard to believe. First, subpersonal experiences hopefully don't have qualitative character, so the experience of producing the string is just the very producing. Suppose that John's competence or parser thus produces the marked string for "M", and the CP takes this producing as its datum; it is almost like John's CP watching the competence-parser producing the whole string. We may assume that upon receiving the word "herself" the competence looks in the dictionary and finds out that "herself" is a kind of word that can play the role in a noun-phrase. It hypotheses the following simple structure:

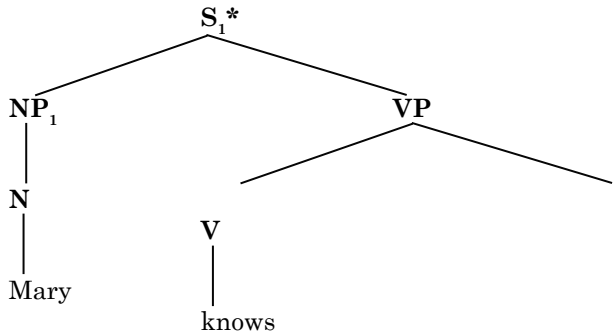


John's CP is watching attentively, we presume on Devitt's behalf. Next come "loves" and "Mary" so the competence happily merges them:



But what kind of information can this give to the CP? Linguistic rules are embodied in the competence, and not in the CP, since it is not a dedicated linguistic processor, but an all-purpose machine. So, mere following the toils of competence does not tell CP much. It has no idea about c-command, and can neither figure out that "Mary" should c-command the rest, nor that it does not do so in the tree. Only the competence has, or rather consists of procedural rules, so only it can decide whether the target string is acceptable. Devitt's hypothesis that "The datum is the experience that the answer is about." doesn't tell us how CP could possibly figure out the verdict.

Things stand even worse with the second example. Assume for simplicity sake, and leaving the technical issues aside, that the parser just follows the order in which it is receiving input, and does not jump much ahead (is not a top-down parser). So, upon receiving "Mary" and "knows" and consulting lexicon, it disposes with the following "conjecture":



What does this tell the CP? How much do you have to know to have even an inkling that a second sentence is expected on the right-hand side? And that an anaphora occurring in that sentence has to follow the rules of c-command? The two finished trees that we copied from the textbook are even more baffling for an ordinary CP. The second option is a non-starter, unless the ordinarist assumes that all people are the linguistic equals of Chomsky.

Let me introduce the third option, the "minimal signal" option, by wondering whether Michael perhaps means something else by his "datum". Maybe, the competence ends up producing only correct strings; if the proposed string is ill-formed it stops, thereby rejecting it, like the "save" function in my word program, that just stops if the file name I want to save contains an illicit letter, thus letting me know that the title is not a correct string of letters. Here is a passage from Devitt that points in this direction; he is talking about "the normal competent speaker":

> If the datum shows that she would have no problem producing or understanding the expression, she is likely to deem it grammatical. If the datum shows that she has a problem, she will diagnose the problem in light of her background theories, linguistic and others, perhaps judging the expression ungrammatical, perhaps judging it grammatical but infelicitous or whatever. Often these judgments will be immediate and unreflective enough to count as intuitions. Even when they do count, they are still laden with such background theory as she acquired in getting her concept of grammaticality. (2006a: 109–10)

On this reading, at the end of its attempt to process the input string the competence signals "No problem" if the string is acceptable, or "I have a problem", if the string is not acceptable (or, even simpler, its (re-)producing the string is the signal translatable as Yes, its having a problem and perhaps aborting the production the signal translatable as No.) So, just by following its toils to the bitter end, the CP can come to know its implicit verdict, in our example the negative one. The third option is just a development of this later alternative: that the answer is just a Yes/No signal in the neural code, translatable by the central processor. To make it more vivid, we can liken it to the red light on the crossing; the pedestrian interprets it as No, and stops if prudent. The ordinarist's view is that this is very little, my view is that this is the core information, the real content of the intuition to be produced at the end of the day.

So, we do have here the voice of competence (albeit a discreet one, liable to be silenced by interferences) and it looks like intuition and feels like intuition. Devitt says explicitly that "(s)omeone who has the relevant competence has ready access to a great deal of data that are to be explained. She does not have to go out and look for data because her competence produces them." (2006: 105) And the whole motivation of Devitt's project was to get rid of immediate access to (deliverance of) competence. Finally, most of the work is done by the answer, our stage no.2, the agreed voice of competence. This happens in two ways. On the one hand, in many cases the stage no.3, empirical testing, does not change the verdict and adds nothing to it. In all these cases, it is the answer (no2) that is in its content identical to intuition (no.4), so its content just is intuition-content. On the other hand, when the verdict is somewhat modified at stage no.3., for instance when it is hedged ("I would never say this, but my kids say it all the time, so I guess it's

OK"), the primary intuitional work is really done by the original verdict (stage no.2).

Consider again Devitt's high demands on the naïve subject. Ann is supposed already to deploy her "folk linguistic concept of grammaticality" to appreciate the connection between this grammaticality and competence in the language. And she knows that she is a competent speaker and so uses herself as a guide to what the competent speaker would do. So she asks herself whether this expression is something she would say and what she would make of it if someone else said it. Her answer is the datum." Does the linguist really need all this from Ann? Devitt himself kindly suggested in correspondence the negative answer. The basic data for the linguist are that Ann would say such-and-such and would mean such-and-such by the given expression. The rest is sociological theory. Remember, when a Chomskyan speaks about one's language, he means one's idiolect. Ann's opinion whether she is a typical representative of a community, whether her English is good English are beside the point. Does Ann have to think that she is competent in the very language (idiolect) she is speaking? The very question sounds ludicrous. Why is this relevant? Because stating the judgment about a sentence in one's idiolect requires *far less theory, if any at all, than reflection about social usage.*

The fourth possibility is that the answer is a relatively articulate verdict, of the kind "No, this is not a well-formed string", again formulated in a neural code. I suppose that Michael rejects it, I would leave it as an open possibility.

To reiterate, I find first the option empty, and the second option, that the datum is experience of simulation itself extremely implausible if taken literally. I tend to agree with the minimal signal option of Yes/No signals, green/red lights in a neural code (and with its subspecies, the more tolerant alternative reading of Michael's "the experience that the answer is about", according to which the successful experience is a green light for the string, the aborted simulation a red light.) It has the advantage of offering a distinct job to competence, and a distinct job to CP, it goes well with poverty of intuitions, their tendency to reduce to Yes/No final verdicts, and it is very parsimonious. However, I have nothing against the articulate signal option, and would be happy if psychologists confirmed it, since this would make more space for competence and help its voice to be heard more easily. End of stage 2.

2.3. *From signal to intuition*

This brings us to the work of CP, our stage 3. Here, two opposite possibilities loom large. Either the empirical theorizing is narrowly linguistic or it is wide, holistic, and could involve language-external, for instance social, affective and other considerations.

So, to summarize, the ordinarist can propose either the wider or the narrower task(s) for the CP. If the task is wider, it is not specifically

linguistic; the linguistic job is done by the competence, and the core of linguistic intuition is really the decision of the competence, it's red or green light. If the task is narrow, it is just translating the very same datum of the competence. Either way, it is the competence that provides the essential core of the resulting intuition: either its linguistic core, on the wide picture, or its total content, on the narrow one, as predicted by The Moderate-Voice-of-Competence view.

How does intuition-capacity develop? We know very little about the improvement of intuition capacity. Some researchers mention that illiterate people did not even understand the questions they asked them: the people just have no idea of what it could be meant by asking whether one would say certain things. Robert Matthews provides a piece of evidence Devitt reports

> This point is nicely illustrated by the following report: "As a graduate student I spent a summer in the Pyrenees (Andorra, Perpignon, etc.) doing field research on the phonology of various dialects of Catalan. Many of our native informants were illiterate peasants. I was forcefully struck how difficult it was to elicit linguistic judgments from them regarding their language, which of course they spoke perfectly well. Just getting the plurals of certain nouns was tough. These folks seemed to be very hard of hearing when it came to hearing the voice of competence! Their difficulty, it seemed, was that their native language was largely transparent to them—they had never thought of it as an object for observation and hence were largely unable to form even the most rudimentary judgments about its character. Catalan speakers with only a modicum of grade school education, by contrast, were good informants, presumably because they had learned through their grammar lessons to think of language as an object with various properties, even if they had no sophisticated knowledge of what those properties might be, theoretically speaking." (Bob Matthews, in correspondence). (2006a: 109 n)

On one reading, the one I prefer, the informants can't distinguish asking about linguistic correctness from asking about pragmatic appropriateness. Maybe, that is, they had difficulties understanding Mathews' questions: is this gentleman asking whether I would say this under some imaginable circumstance (e.g. when drunk, or joking), or whether it would be an appropriate statement to make, or something else? It is hard to believe that in their normal life they do accept systematically ill-formed utterances: in a village, a person with linguistic deficit might be severely ridiculed, and the ridicule seems to speak in favor of other villagers having very definite intuitions to the effect that something is badly wrong with the poor person's way of talking.

Let me conclude with a few short remarks about the list of problems in waiting, connected to the notion of basic competence, as developed by Chomsky. How much of its power is due to the innate structure, how much to empirical learning? Let me, with the majority, call the descriptive-explanatory view that stresses empirical learning "empiricism", and reserve the term "rationalism" for the opposite descriptive-explanatory view, that experience plays no essential role in the process, beyond mere triggering or prompting. (I shall later be calling

the corresponding normative views, concerning justification "aposteriorism" and "apriorism" to avoid the confusion between the normative and the descriptive-explanatory.) The two dynamic processes, relevant here, are first, acquiring the language, and second, improving one's intuition capacity, and they both constitute an important area of research. Both seem to involve straightforward learning. Even on the most mainstream Chomskyan view, which Michael of course rejects, the child's language module needs external information to fix the parameters characterizing the language of the surroundings. In later stages, the child will learn more and more sophisticated facts about the language that is going to become her mother-tongue. Does this imply (descriptive) empiricism about intuition? Only a bit of it: the scaffolding is given, but some information is learned from surroundings. And it might be interesting do discern the fine structure…

In my paper "Intuitions: the discreet voice of competence" (2006), I have argued that our innate endowment might explain at least the very origin of the basic intuition-capacity and the initial stages of the formation of our intuition-states with their contents, but that nativism should be restricted to the origin of the system and to the relatively initial stages of processing. Does the incontestable fact that people's intuitions do develop, and that sophisticated, well-trained people have much richer intuitions speak in favor of empiricism? Again, only to some extent.

Let me conclude this part of the story by stressing the passion for explanation, omnipresent in the debates about linguistic intuitions. This passionate "explanationism" has been in the past characterizing the debate about all sorts of intuitions, in the work of classics, like Plato, Descartes, Kant, and the early twentieth-century philosophers like Russell and Husserl; it is a pity that present-day debate on intuitions in general is much less marked by it.

How do the stages of intuition-production fit with the theoretical work in linguistics? The main way a linguist arrives at her theory is nowadays by focusing upon speaker's intuitions. We may add this development to our story about stages. Remember, the two last stages were the explicit intuition and the report. We now add three more stages. The first is putting intuitions together, and streamlining them (weeding out contradictions, mistakes and the like); it corresponds to building a narrow reflective equilibrium in the case of other kinds of intuitions. Then comes narrow grammatical-linguistic theorizing, and finally balancing the results with the input from other disciplines: psycholinguistics, neurology, possibly evolutionary biology, and science that could help. The result, the Final theory, would be something like a wide reflective equilibrium, encompassing intuitional and other data, input from other sciences, and linguist's theorizing.

4	5	6	7	8	9
→ intuition	→ report	→ organizing the reports streamlining them	→ narrow reflective equilibrium	→ input from other sciences	→ the Final Theory = wide reflective equilibrium

So much about intuition. We now turn to the other set of questions, the ones concerning the metaphysical status of linguistic entities.

3. *Validity of intuitions: Production-and response-dependence*

3.1. *SLEs are production-and-response-dependent*

The idea proposed and briefly defended above that intuition is just another window upon the linguistic competence suggests that language production, understanding and judging go together. It helps to make our view of language abilities more unified and systematic. Moreover, such a unified picture suggests also a metaphysical framework for locating language. Here, the question we want to answer is: where is language situated in the general order of things? In the mind, in the world, or somewhere in-between, straddling the divide. Call this question, following the usual usage in metaphysics, the location problem for language. Here is how it arises.

Commonsensically, the way the folk hears and sees the linguistic tokens seem to suggest that the language is in the outside world, it is "the spoken and written stuff", with no narrower specification. It is what was later called E-language by Chomsky. But the folk wisdom has been put in doubt, since rather early times, at least since the second century AD. Sextus Empiricus in his treatise *Against the Grammarians* questions the existence of language: for him, the availability of sound, syllable, word and sentence are highly dubious. Sounds are questionable since we don't have a clear principle of identification that would tell us, for instance, if a diphthong is one sound or two or whether a hardly audible "r" in a syllable is a sound or not (I, 117 ff.). Words and parts of the sentence, inherit the dubiousness; "if the aggregate of the parts of the sentence is conceived to be a sentence, then because the aggregation is nothing apart from the parts aggregated, just as distance is nothing apart from the objects which are distant, the sentence of which any parts shall be conceived will not be anything. And when the whole sentence is nothing, neither will any parts of it exist. I. 135, p. 81). The conclusion is that "neither does the sentence exist" (I, 137). Many contemporary linguists and philosophers of language would agree with Sextus: in their view, already phonology shows that there is a problem: the phonetics-phonology interface is complex, and what is heard depends heavily on context, and more importantly on expectations and habits. This internal, mental component has become even more central for the linguists with the advent of Chomskyan linguistics. The question now looms large: where should we locate the language?

Let me detail the problem a bit. Chomsky's insistence upon the importance of I-language has generated the temptation of radical internalizing motivated by the wish completely to get rid of the first, external component, that is seen as too contingent, "non-linguistic" and unsystematic to be worthy of being called "language" in the strict sense at all. "Language, as far as I can tell, is *all* construction" writes Jackendoff (1992: 104, emphasis by Jackendoff), echoing Sextus after almost two millennia. He also initiated the comparison between "hearing" the linguistic items and "seeing" visual illusions. The invidious comparison is later developed by G. Rey (2005) a fervent defender of the picture of language as merely internal, even "imagined" and also used by Isaac 2008: 24). Phonemes, their aggregates (morphemes, internal sentence-strings), syntactic trees and other "standard linguistic entities" (SLEs to use the abbreviation preferred by G. Rey) are in the head, and they are the whole of language. Language is I-language, E-language is out.

Another example is the blank screen metaphor offered by John Collins: E-language is like the blank screen onto which linguistic items are being projected by our mind. It sounds there is literally nothing out there. Take the book with complete works of Shakespeare, and take an empty book, just blank pages like movie screen; it can't be that in the case of his complete works there is nothing (written in a language) there, that we project the text upon blank pages. Collins has responded in the correspondence that he attacked only extreme objectivism, without really wanting to deny that there is E-language outside. But saying, as he does, that "the tokens are there alright, but the properties are simply projected onto them" bring the blank screen back in; the tokens don't have any linguistically interesting properties.

Philosophers are sometimes tempted to extremes: in this case to extreme internalization of matters linguistic. The temptation is to halve the reality of language. Commonsensically, the language is outside, it's the spoken and written stuff. Comes Chomskyan linguistics and detects another half, the internal one, which turns out to be quite important. This generates the radical internalizing temptation to get rid of the first half. But such a Sextan line has to be resisted, on the pain of linguistic nihilism. Commonsense entities should not be liquidated unless it is really necessary. Widespread illusions are not to be postulated unless this is absolutely unavoidable. Notice that utterances (speech, books, conversations) are (still) part of explanandum in linguistics, so E-language is (still) part of the linguistic enterprise. The E-language is needed first as crucial input: you need utterances that are acceptable or not. Note the importance of the gathering of data, and the fanatical attention to the details in the data that characterizes great linguists. The strong internalizing move tends to make nonsense out of this importance, and turns linguistics into a rather strange enterprise of non-accounting for anything at all.

Devitt prefers a strong externalizing move. Language is just a piece of non-mental reality, it's physical signs and body-involving practices,

and nothing mental at all. He contrasts linguistic reality to mental reality.

Second, the theory building proceeds by accounting for the acceptability of these E-language items. The impressive machinery of UG is consistently being proposed as an account of reactions of native speakers to utterances proposed or thought of (when the linguist is using herself as the linguistic guinea-pig) limited variability of data. The data cannot vary indefinitely and in a chaotic fashion. Especially if simplicity is such a prime ambition: if all of the grammar is to be captured in a few very simple and very general principles, then the data should be highly organized: all the data should be amenable to the principles. Why is the idea of parameter setting so impressive: because it detects a deep, very general uniformity, and constraints variation with an iron hand!

Finally, note the contrast between folk-superstition and folk view of language as being out there in the world. The first ascribes all sorts of crazy powers to presumed witches, which the women in question just don't have. Folk-views on language don't ascribe any crazy stuff to sentences. They just fail to add sophisticated stuff, that's hardly an error. So, we need a way to preserve the folk or commonsense view that utterances *are* language.

On the other hand, the path back to the Eden of commonsensical firm objectivism is also closed, since too much of linguistic material is clearly located strictly in human cognitive apparatus. So, why not go part of the way with the folk, give unto complete works of Shakespeare what belongs unto them, and give to the mind what belongs unto it. The answer I favor is therefore that the two halves, the internal and the external, E-language and I-language, have to be kept together. This is compatible with a wide range of views, from eliminativism (linguistic items are only projected)1 and quasi-eliminativism about them 2, to various views that stress speaker's intention as basic for linguistic product(ion)s to the classical Chomskyan stance according to which language is essentially a mental thing, and only accidentally has an external manifestation. So, how do we choose? I think that the best way is the most direct one: language is relational and response-dependent, But the answer needs a lot of developing and defending: pointing out that it is in the middle between the two extremes would have been able to recommend it in times of Aristotle, when the virtue was considered to lay in the middle.

In brief, linguistic reality of a language, say French, is pretty much determined by the psychological reality of its speakers. The mental dispositions and states of native French speakers, determine in the last instance, the structure (syntactic, semantic and pragmatic) or our language. Capacity of recognizing the correct forms, crucial for intuitions, is part of this constitutive structure. A sound-string is a correct sentence of a language partly because a competent language speaker would recognize it as such. It is obvious that p is a correct sentence of

the language iff a normal thinker would find *p* clear and compelling. In this sense, the validity of intuitions is ultimately explained by the efficiency of the wider collective psychological make-up of which they are part. The order of determination seems to go from the mind to the world. Language is in-between the mind and the world, straddling the divide. Let us look at more detail.

Start with the level of phonology. Why is a certain stream of sounds, or a series of inscriptions a token of "M"? Because John would have intuition that it is, and hear it and read it as a token of "M". To pass further to graphemes, why are the following two very differently looking strings in fact tokens of the same Croatian sentence?

(L) Svaki od dva kandidata očekivao je da će pobijediti onog drugog.
(C) Сваки од два кандидата очекивао је да ће побиједити оног другог.

Because a Croatian reader knowledgeable of the Cyrillic script would see them so, and, if need arose would write his thought in the form of (C). I just did it on my computer, and I count as a Croat with a command of Cyrillic.

So, the linguistic properties of the sentence seem to be dependent on the production and on the responses of the speaker in charge. Let me then introduce a new term, with a play of words to boot. Call this hypothetical dependence *"production- and response-dependence", PR-dependence for short*. You might take the "PR" as a pun, reminding one of "public relations", which is not bad: a more psychologically minded reader might read the term strictly as referring to dependence on individual speaker-hearer, a more sociologically minded one as pointing to further dependencies. So, a simple relational answer to our metaphysical location problem would be that linguistic entities are production- and -response dependent. In the form of a slogan, *SLEs are PR-dependent*. (Thanks go to Michael who in discussion warned me that pure response-dependence is very, very implausible; he would not accept the PR-dependence either.

Let me explain the slogan and argue for it in more detail. The slogan claims the following about any given SLE:

> SLE characterizes a given linguistic (external) token *t* of L iff a normal speaker-hearer of L would accept *t* and would produce *t* if suitably prompted.

And, the left-hand side determines the right hand one: SLE is a property of a given linguistic (external) token t because the normal speaker of L would do as specified. With the advent of Chomskian linguistics it has become a commonplace that competence in a way *dictates* what counts as the correct, well-formed sentence of a given language. Let us pass to illustrations and explanations.

Start with the lowest level. The commonalities between different tokens of the same sentence like our (L) and (C), but also of tokens variously articulated, are recognizable only through the detour through the

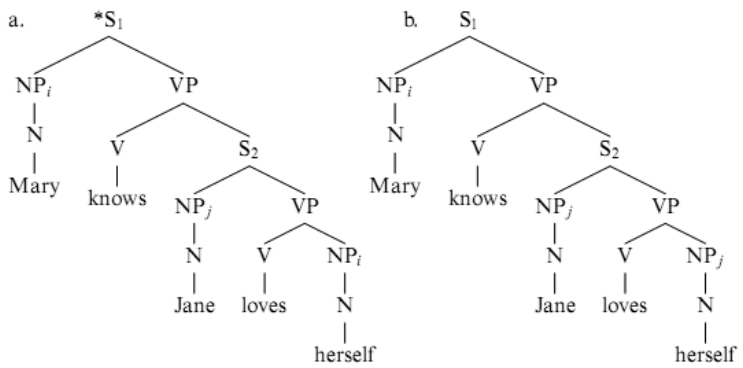
response of the hearer, call it response-detour. The commonalities are also *explainable* only in terms of the response-detour. Therefore, by inference to the best explanation, the fundamental phonological (-graphematic) properties of the strings are not mind-independent properties. An external token string has linguistic properties because it would be interpreted as having them by the normal language-hearer, and would be produced by a process that would form it respecting the nature of these properties. So, they are best seen as PR-dependent properties

The same is valid for fundamental grammatical (-semantic) properties of the strings, bringing in our discussion of intuitions. We focus here upon the first family, the syntactic one.

In researching a given language the linguist has to do with speaker-hearer, the person who not only produces a string, but also reacts to the string itself, so that her reaction, in the form of her intuition and report, offer a glimpse into the structure of her I-language. Return for the moment to the sentence

(M) Mary knows that Jane loves herself.

and to our two trees.



Now, why does the sentence "M" say that Jane loves herself, Jane, and that Mary knows this? Why not that Jane loves Mary; that might have been a happier situation. Well because John, the speaker-hearer of the language, has intuitions that can be systematized by the two trees, and has, presumably, something in his cognitive apparatus that is roughly isomorphic to them (either explicit representations, or a sequence of processing moves). If he heard "M" systematically in the "Jane-loves-Mary-who-knows-about-it" way, and produced it to indicate this, then the sentence would have had a different (a linguistically "illegal") syntactic structure represented by the left-hand monster tree. Moreover, John would himself produce "M" to mean that Jane loves herself, Jane, and would thus testify to the fidelity of the "legal" analysis. The two combine in simple cases. (The exceptions and complications concern the "impure" intuition we talked about earlier, for instance our hero Ina who allows ungrammatical strings out of her love for the non-na-

tive partner, but does not produce them herself. The other, and indeed famous example is the opposite one, studied by Labov: the snobbish reaction of not officially allowing utterance-tokens of the kind one produces oneself).

So, the property of the given utterance of "M" of being a token of a sentence with the c-command pattern that makes it into a correct piece of John's language, looks like a relational property, that the token has in virtue of playing a (potential or actual) role in awakening John's responses, including the intuitional ones, and of being potentially produced by John's linguistic apparatus as a correct item. So, in general it seems that the relevant syntactic properties are relational properties, that the physical type sentences posses in virtue of being related to psycho-grammar (in terms of being produced/-able and being recognizable/parsable/ as such), i.e. PR-dependent properties. The production can be taken in one of the two ways, and I will remain officially neutral between them. For the representationalists about SLEs, respecting a mental phrase marker is to be taken literally: when John is producing the sentence "M", his mental apparatus constructs the marker and then organizes the sentence in accordance to the representation produced (merging and moving its elements with one eye on the represented paradigm). A non-representationalist will settle for less: the apparatus functions as if it is implementing the phrase marker, it follows the rules without representing them, the way planet's follow Kepler's laws, blindly but reliably. In both cases the best choice is our slogan: SLEs are PR-dependent.

But isn't the tree still merely projected onto the sentence-token by John's mind, a strict internalist might ask. In answer, we should think of degrees of projection versus guidance. The first and clearest is total projection upon a blank screen, like the movie of tv-screen. The next is moderate projection like in the Rorshach test, where the external shapes play an auxiliary suggesting role. The third is a mix, combining minimal projection with moderate guidance. Take projecting depth or perspective into a realistic-style painting of an imaginary landscape. The painter has done most of the work, her hand guides the eye of the observer; still, an observer with no experience with paintings or photographs will have trouble noticing depth. Consider how the three cases reflect on the study of the relevant phenomena. First, nobody studies blank screens in order to understand various movies, but, second, psychiatrists propose better or worse test, without normally trying to account for every point in the drawing within a systematic theory. And finally, concerning the recovering of depth-information, the typical art historian pays maximal attention to almost every feature of the realistic-style painting. The painting shows with great exactness what *the painter had in mind.* Linguists come *closest to the third case*, with endless, patient search for examples and counterexamples, pairs of very similarly looking written sentences, one of which is acceptable

and other not. So, the right mix for language is *minimal projection and maximal guidance.* The corpus shows with great exactness what *the speaker had in mind,* sub-personally, of course. And indeed, a volume of prose or drama in a given language hardly reminds one of blank screen; a post-modernist might see in it a kind of Rorschach, but the scientific linguistic attention to detail points to the opposite end, the minimal projection**,** exactly as the PR-dependence picture would predict.

Let me conclude by mentioning several interesting issues related to the narrow response-dependence, i.e. dependence on the hearer. Suppose that I want to say that Mary is aware of Jane's love for her, but, having previously had a nice portion of my favorite brandy, I produce a token of "Mary knows that Jane loves herself". "You claim she knows that Jane is narcissistic?" a colleague asks. Well, this is what I have said, and here it is the hearer, the colleague, and her response, that decides. The same goes for meaning, in particular public meaning vs. speaker's meaning. The speaker thinks that "promiscuous" is a very posh word for "promising" and says: "I like promiscuous candidates." He meant "I like promising candidates", since this is how he composes (produces) and hears (responds to) his sentence. Unfortunately, the public meaning is the other, slightly awkward one. Why? Well, my dear speaker, because most people would hear your sentence as talking about candidates who change their partners a bit too often (for some standards). Looks like the public meaning is pretty much narrowly response-dependent, defined by how people would decode the utterance. Now, if (and I say *if*) the public meaning has primacy, then the hearer is privileged in relation to the speaker, and we have strict, narrow response-dependence.

On the other hand, in the case of ambiguity, as Devitt has warned me in a pleasant discussion in the IUC in Dubrovnik, the speaker seems to have the upper hand. If I say, "Visiting relatives can be boring", meaning the ritual of visiting them, and you take it the other way around, it is me who is the authority. So, dependence on production remains, along with the one on response. SLEs are PR-dependent.

3.2. *Sketching the big picture: Intuitions and the maximalist response-dependentist view*

How does the view just sketched, fit with the rest of our world picture? Are there other PR-dependent properties? What about the moral properties like *right*, *just* and *fair?* Many people swear these are fully objective, others see them as "social constructions" or projections. But there is a middle way One can say that an action was just if it would strike an impartial observer as being such (or, impartial observers would agree to this effect, under one or other guise, including veil of ignorance and similar devices). Kantians would insist on the production: the action was right only if it was produced with the right intention, the one of doing what is morally required by the situation. If we combine the two, we get the PR-dependence:

> An action a is morally right iff it has been produced with the intention of doing what is right and is perceived by impartial observers as right.

This line has not been explored, and I rest content with just mentioning it here. (I got the idea of the moral-linguistic parallel in a conversation with Nenad Smokrović, so thanks go to him). Obviously, the language, if to some extent mind-dependent, is certainly not alone in being such. For pure response-dependence take secondary qualities like *bitter* and *red*. They are hardly production-dependent, but it is arguable that they are response-dependent, and I have argued for the view in several papers. We may introduce the term "generalized response-dependence", to cover items like SLEs that are both production-and-response-dependent and those that are response-dependent in strict sense. This would yield the following division:

response-dependent in general

↙ ↘

narrowly production-and-response-dependent

response-dependent = PR-dependent

Now, all the examples (bitter, red, SLEs) are cognitive, as opposed to moral-practical. We might add other cognitive examples of the same high level, for instance being obvious; it is a property typically accompanying intuition-contents. When is it obvious that something is the case, e.g. that our (M*) is an incorrect sentence:

> It is obvious that p iff a normal thinker would find p irresistible, clear and compelling.

And, the left-hand side determines the right hand one: it is obvious that (M*) is incorrect because the normal speaker of John's language would find it compelling that it is so. Our next non-practical example are expressive properties, like e.g. *happy-looking*. A person is happy looking because it looks happy to her surrounding, more officially because it would look happy to the relevant group of normal observers. What about individual emotional qualities like *frightful* or *sad* as predicated to objects and situations? They might be response-dependent as well: a scene is frightful because it would frighten the normal observer. Our last non-practical example is much more philosophical, aesthetic properties like *good-looking, beautiful* and *ugly*. Since Hume's times they have been characterized by many philosophers in response-dependentist terms.

Passing now to the minefield of other practical candidates for response-dependence beside the moral qualities, let me start with the least problematic candidates. Take individual attraction and repulsion, with qualities like *attractive, disgusting, sexy* and the like; they are probably the least problematic candidates for a response-dependentist

account. Social properties like *inviting* and *offensive* follow suit. Like in the case of language I am in favor of a response-dependentist view. Similarly with the "significance" properties, like *meaningful* or *meaningless*, particularly for items like a meaningful life.

This is then the first sketch of the big picture: many, perhaps all humanly interesting properties are response-dependent in the very general sense indicated (involving production-dependence besides the narrow dependence on the response of the hearer); language is here joined by its peers within a huge family of properties.

4. *Conclusion: Explaining intuitions*

Let me recapitulate. In the first part I have defended a moderate "voice-of-competence" view, the view that there are intuitions-dispositions and judgments, which form a distinct group of phenomena, and there is the intuition-capacity, the capacity to use our linguistic competencies in an off-line fashion. It is the voice of competence, most often a discreet one. Intuitional data are thus the minimal "products" of tentative linguistic production and primarily not their opinions about the data. The data involve no theory and very little proto-theory. Although there might be admixtures of guesswork in the conscious production of data, these are routinely weaned out by linguists. In contrast, ordinarism denies that seemingly linguistic judgments form a significant epistemic kind, intuitions, and that there is a distinct capacity producing them. I agree with ordinarist about referentialism: intuitions are concerned with their external objects, the domain of items and facts. I agree about the importance of explanationism in contrast to quietism: items traditionally described as intuitions require an explanation of having and reliability, if possible a causal one. However, I have defended against ordinarists the competence-centered option, that is also referentialist and explanationist, and proposed a sketch of explanation, featuring the hypothetical stages of intuition-generation and the less hypothetical ones of reflection about intuitions.

The second part develops a very rough sketch of what in the reality validates our linguistic intuitions. For this, it turns briefly to the metaphysics of language, placing it in-between two extremes, the purely psychological Chomskyan one and the purely extra-mental, the Devitt's one. With the advent of Chomskyan linguistics the old question of where language is situated in the general order of things has been partly answered by stressing the role of the mind. The order of determination seems to go from the mind to the world. The view defended here agrees but adds and stresses that language is in-between the mind and the world, straddling the divide, and that a sound-string is a correct sentence of a language because a competent language speaker would recognize it as such. To put it in a more technical sounding terminology, language is production- and response-dependent, and being response-dependent goes together with color, moral, epistemic and aesthetic

value and with emotional properties. The two parts offer a sketch of a relatively encompassing view of language and linguistic intuitions, combining epistemological and metaphysical topics.

References

Collins, J. 2008. *Chomsky: a guide for the perplexed*. London: Continuum.

Devitt, M. 2005. "There Is *No A priori*." In E. Sosa and M. Steup (eds.). *Contemporary Debates in Epistemology*. Oxford: Blackwell Publishers.

Devitt, M. 2006a. *The Ignorance of Language*. Oxford: Clarendon Press.

Devitt, M. 2006b. "Intuitions in Linguistics." *The British Journal for the Philosophy of Science* 57: 481–513.

Devitt, M. 2006c. "Defending Ignorance of Language, Responses to Dubrovnik Papers." *Croatian Journal of Philosophy* 6: 571–605.

Devitt, M. 2012. "The Role of Intuitions." In G. Russell and D. Graff Fara (eds.). *The Routledge Companion to Philosophy of Language*. London: Routledge: 554–563.

Ewing, A. C. 1971. *Fundamental Questions of Philosophy*. London: Routledge.

Isac, D. and Reiss, Ch. 2008. *I-Language an Introduction to Linguistics as Cognitive Science*. Oxford: Oxford University Press.

Jackendoff, R. 1992. *Languages of the Mind*. Cambridge: MIT Press.

Miščević, N. 2006. "Intuitions: The discrete voice of competence." *Croatian Journal of Philosophy* 6: 523–548.

Moore, G. E. 1991. *The Elements of Ethics*. Edited and with an introduction by Tom Regan. Philadelphia: Temple University Press.

Rey, G. 2005. "Mind, Intentionality and Inexistence." *Croatian Journal of Philosophy* 5: 389–415.

Sextus Empiricus. 1949. *Against The Professors*, *Book One, Against the Grammarians*. Cambridge: Loeb.

Croatian Journal of Philosophy
Vol. XVIII, No. 53, 2018

Intuitions: Rijeka Response to Nenad Miščević

MICHAEL DEVITT
The Graduate Center, the City University of New York, New York, USA

This paper is a response to Nenad Miščević's "Reply to Michael Devitt", the latest in an exchange on the source of linguistic intuitions. Miščević defends a modified version ("MoVoC") of the received view that these intuitions are the product of a linguistic competence. I have earlier rejected all versions of the received view urging instead that intuitions are, like perceptual judgments, empirical theory-laden central-processor responses to phenomena. (1) I emphasize here, against Miščević, that this claim about a speaker's intuitions about strings is not to be conflated with a claim about her understanding of strings. (2) I develop my claim, addressed by Miščević, that MoVoC is implausible in three ways. But these are not the main problems for MoVoC. For further discussion of those, see Jutronić's paper in this volume.

Keywords: Voice of Competence, Modest Explanation, linguistic understanding, visual analogy, developmental evidence.

Nenad Miščević (2006, 2009, 2012, 2014a, b), Dunja Jutronić (2012, 2014), and I (2006c, 2014) have been in an exchange about linguistic intuitions for more than a decade. I have found this very productive but I suspect that we have now reached the point of diminishing return. So I shall be brief.

1. *VoC ("The Voice of Competence")*

I claim that VoC is the received view of intuitions in linguistics. Consider the intuitive judgments that

(1) John seems to Bill to want to help himself

is a grammatical/acceptable sentence, and that in it 'himself' co-refers with 'John'. VoC is the view that these intuitions are the product of a linguistic competence residing in a sub-central module of the mind. I describe VoC as the view that linguistic competence, all on its own,

> provides information about the linguistic facts… So these judgments are not arrived at by the sort of empirical investigation that judgments about the world usually require. Rather, a speaker has a privileged access to facts about the language, facts captured by the intuitions, simply in virtue of being competent… (2006a: 96; 2006b: 483–4).

On this view, competence not only plays the dominant role in *linguistic usage*, it also provides informational content to *metalinguistic intuitions*. Those intuitions are indeed, "noise" aside, the voice of competence. That is why they are reliable.

Miščević holds to a "modified" version of VoC, "MoVoC". I reject all versions of VoC partly because I think I have a better view, "The Modest Explanation", which unfortunately has been labeled "ordinarism".

2. *The Modest Explanation or "Ordinarism"*

The Modest Explanation of linguistic intuitions arises from a view of intuitive judgments in general. I argue that intuitions are empirical central-processor responses to phenomena, differing from many other such responses only in being fairly immediate and unreflective, based on little if any conscious reasoning. These judgments are theory-laden in the way observation judgments are in general. Related to this "we should trust a person's intuitions…to the degree that we have confidence in her empirically based expertise about the kinds under investigation" (2006a: 104; 2006b: 492). Sometimes the folk are as trustworthy as anyone in an area but where there are experts, particularly scientists, the folk are not. So we should prefer the paleontologist's intuitions about old bones, the physicist's, about certain physical facts, the psychologist's, about certain cognitive phenomena, the art expert, about a sixth-century Greek marble statue, and Vic Braden about tennis serves (2006a: 104–5; 2006: 492–3). And when it comes to linguistics, "the intuitions that linguistics should mostly rely on are those of the linguists themselves because the linguists are the most expert" (2006a: 111; 2006b: 499).

Just as the paleontologist, the art expert, and Vic Braden, immediately recognize the relevant property in their cases, so too does the speaker in easy linguistic cases. Consider the strings, 'responded the quickly speaker' and 'the speaker responded quickly'. I have pointed out that "the speaker is likely to recognize immediately, without reflecting…, that the former word salad is unacceptable and the latter simple sentence is acceptable" (2010a: 255). In my paper, "Linguistic Intuitions: In Defense of 'Ordinarism'" (2014), I emphasize that these "linguistic intuitions are *perceptual* judgments…as immediate as those of the art expert and Braden, without the conscious and deliberate exercise of her competence" (2014: 14).

In his "Reply to Michael Devitt", Miščević says he agrees "completely" with this claim (2014a: 25).[1] But he continues on as if my claim is

[1] Hereinafter, all citations of Miščević and me are of these 2014 papers unless specified otherwise.

about a speaker's *understanding of* strings: that understanding is what he takes to be like percepts. Yet my actual claim is not this quite uncontroversial one but rather that *intuitions about* strings are like percepts. The conflation of this crucial distinction between understanding a string and having an intuition about it runs right through Miščević's discussion. This is a common conflation among defenders of VoC, as I have pointed out (2010b: 839). That is my first point in this response to Miščević.

Let me give just one other example of the conflation. In response to my claim,

> To say that a speaker may perceive that a string has a certain syntactic property without *a conscious and deliberate* exercise of her competence… is not to say that her competence is not involved in her perception (p. 14),

Miščević responds: "if the competence is involved, as the just quoted passage suggests, why is VoC on the wrong track?" (p. 26). And the answer is: because the competence is involved in understanding the string not in providing the informative content of the intuition, as VoC requires.

3. *Criticism of VoC (MoVoC)*

So why should we prefer ordinarism to VoC? I quoted (p. 11) the following summary of my reasons:

> The main problems with it are, first, that, to my knowledge, it has never been stated in the sort of detail that could make it a real theory of the source of intuitions. Just *how* do the allegedly embodied principles yield the intuitions? We need more than a hand wave in answer. Second, again to my knowledge, no argument has ever been given for VoC until Georges Rey's recent attempt (2013) which, I argue (2013), fails. Third, given what else we know about the mind, it is unlikely that VoC could be developed into a theory that we would have good reason to believe (2015: 37).

I went on to look critically at Miščević's MoVoC in light of these problems (pp. 11–12). So too did Jutronić (2014). Miščević has responded to her (2014b). I leave it to her to pursue the matter (this volume).

Apart from these three main problems, I also summarized three other implausibilities of VoC (p. 12). Miščević addresses these in his reply to me and so I shall develop them in more detail here.

(i) There are clearly lots of linguistic facts about which ordinary speakers have few if any intuitions: facts about heads, c-command, and so on. Why is that? "If our competence…speaks to us at all, how come it *says so little*?" (2006a: 101; 2006b: 489). We wonder what account of the causal route from embodied rules (and principles) to intuitions could account for this. In his reply, Miščević emphasizes how many intuitions about grammaticality and binding hearers have (p. 23). They do indeed have many tokens of these types of intuitions. But my point was about types not tokens: How come competence does not deliver intuitions about

many other types of grammatical facts, like about c-command facts?

(ii) Chomsky has, in effect, found support for VoC in an analogy with the intuitions yielded by the visual system (1965, pp. 8–9; 2000, p. 125). Others have followed him in this (Rey 2006, pp. 563–8; Collins 2007, p. 421; Fitzgerald 2010: 134–42). I have rejected the analogy (2006a: 112–3; 2006b: 500–1; 2010b: 850–2). Indeed, I argue that there would be a disanalogy between the intuitions provided by the language faculty and by perceptual modules which undermines VoC. Miščević doesn't "get it. Why would the MoVoC proposal be committed to any such disanalogy?" (p. 23) Here is the answer that I cited:

> According to the standard explanation, the language module delivers syntactic and semantic information about expressions to the central processor. If it did this it would be disanalogous to perceptual modules…. For, if [the language module delivers syntactic and semantic information about expressions to the central processor], the central processor would have direct access to information that the language module allegedly uses to fulfill its task of processing language. But nobody supposes that the central processor has direct access to analogous information used by perceptual modules to fulfill their processing tasks (2006a: 114; 2006b: 503).

(iii)

> Developmental evidence suggests that the ability to speak a language and the ability to have intuitions about the language are quite distinct, the former being acquired in early childhood, the latter, in middle childhood as part of a *general* cognitive development. (2015: 37)

Miščević doubts that these developmental stages are really distinct, citing a "proposal that ties the acquisition of full competence to a later stage" (p. 24). This proposal talks of

> language users who might be described as 'not fully competent', such as very young children, second language learners, or aphasics who have lost access to part of their language competence. Such language users may resort to simplified strategies or heuristics for sentence processing (Ingram 2007: 18).

The evidence I cite (2010b: 853 n. 27) distinguishes two stages of cognitive development, early childhood up to around 3, and middle childhood from around 4 to 8. It is alleged that syntactic competence is achieved in the early stage, metalinguistic intuitions, in the middle one. This could be quite consistent with what Ingram claims; in particular, "very young children" may be ones who have not completed early childhood.

I emphasize that implausibilities (i) to (iii) are not the main problems for VoC and MoVoc. For further discussion of those main problems I direct the reader to Jutronić's paper in this volume.

References

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chomsky, N. 2000. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Collins, J. 2007. "Review of Devitt 2006a." *Mind* 116: 416–23.

Devitt, M. 2006a. *Ignorance of Language*. Oxford: Clarendon Press.

____2006b. "Intuitions in Linguistics." *British Journal for the Philosophy of Science* 57: 481–513.

____2006c. "Defending *Ignorance of Language*: Responses to the Dubrovnik papers." *Croatian Journal of Philosophy* 6: 571–606.

____2010a. "What 'Intuitions' are Linguistic Evidence?" *Erkenntnis* 73: 251–64.

____2010b. "Linguistic Intuitions Revisited." *British Journal for the Philosophy of Science* 61: 833–65.

____2013. "Linguistic Intuitions are not 'the Voice of Competence'." In Haug 2013: 268–93.

____2014. "Linguistic Intuitions: In Defense of 'Ordinarism'." *European Journal of Analytic Philosophy* 10 (2): 7–20.

____2015. "Testing Theories of Reference". In J. Haukioja (ed.). *Advances in Experimental Philosophy of Language*. London: Bloomsbury Academic: 31–63.

Fitzgerald, G. 2010. "Linguistic Intuitions." *British Journal for the Philosophy of Science* 61: 123–60.

Haug, M. (ed.). 2013. *Philosophical Methodology: The Armchair or the Laboratory?* London: Routledge.

Ingram, J. C. L. 2007. *Neurolinguistics. An Introduction to Spoken Language Processing and its Disorders*. Cambridge: Cambridge University Press.

Jutronić, D. 2012. "Je li jezična sposobnost izvor jezičnih intuicija (Is languge competence the source of linguistic intuitions)." In S. Prijić-Samaržija and P. Bojanić (eds.). *Nenad Miščević – Sva lica filozofije (Nenad Miščević – all faces of philosophy*). Beograd: Institut za filozofiju i društvenu teoriju: 129–143.

Jutronić, D. 2014. "Which are the Data that Competence Provides for Linguistic Intuitions?" *European Journal of Analytic Philosophy* 10 (2): 119–43.

Miščević. N. 2006. "Intuitions: The Discrete Voice of Competence". *Croatian Journal of Philosophy* 6: 523–548.

____2009. "Competent Voices: A Theory of Intuitions". http://oddelki.ff.uni-mb.si/filozofija/en/festschrift

____2012. "Odgovor Dunji Jutronić (Answer to Dunja Jutronić)". In S. Prijić-Samaržija and P. Bojanić (eds.). *Nenad Miščević – Sva lica filozofije (Nenad Miščević – all faces of philosophy*). Beograd: Institut za filozofiju i društvenu teoriju: 194–198.

____2014a. "Reply to Michael Devitt." *European Journal of Analytic Philosophy* 10 (2): 21–30.

____2014b. "Reply to Dunja Jutronić." *European Journal of Analytic Philosophy* 10 (2): 145–53.

Rey, G. 2006. "Conventions, Intuitions and Linguistic Inexistents: A Reply to Devitt." *Croatian Journal of Philosophy* 6: 549–69.

____2013. "The Possibility of a Naturalistic Cartesianism Regarding Intuitions and Introspection." In Haug 2013: 243–67.

Croatian Journal of Philosophy
Vol. XVIII, No. 53, 2018

Intuitions Once Again! Object-level vs. Meta-level

DUNJA JUTRONIĆ
University of Split, Split, Croatia

Firstly, I present some of my most important answers to Miščević's objections to my 2014 paper which I fully disagree with. Secondly and more importantly, I point out that there is a possible confusion or misunderstanding about the distinction between the object-level (sentence produced) and meta-level (sentence judged). I argue that competentionalist actually conflates object and meta levels and show the final consequences of such a conflation. The ordinarist firmly believes that there should be a separation between the object-level and meta-level and provides the explanation for this. Finally, I briefly comment on the so-called 'Route Question', the path from the underlying competence to the central processor and argue that competentionalist cannot provide an explanation for it. The hope is that this discussion brings us closer to understanding the difference between the two opposing views.

Keywords: Linguistic intuitions, competentionalism, ordinarism, meta-level, object-level, the route question.

1. *Introduction*

This paper came out of a mini conference with Michael Devitt that was held in Rijeka under the title 'Linguistic Intuitions and Natural Kinds' in April 2017 organized by the Croatian Society for Analytic Philosophy. The part on linguistic intuitions was meant to be a continuation of Devitt's, Miščević's and Jutronić's ongoing discussion on linguistic intuitions[1] and here I present some of my most important answers to Miščević's objections to my 2014 paper. However, my main aim is to try to advance the discussion which, in my opinion, is coming to an un-

[1] See Jutronić (2012, 2014); Miščević (2012, 2014); Devitt (2014). The literature on linguistics intuitions has proliferated greatly. For more references see Jutronić (2014) and Schindler (Forthcoming).

profitable stand still. My hope is that this discussion brings us closer to understanding the difference between the two opposing views!

2. *Answers to Miščević*

Let me start with a simple example. Say a speaker is presented with the following two sentences:

They want to be teachers.
*They want to be teacher.

If asked, the native speaker answers that only the first sentence is OK, or that this is a sentence of his language. His answer is the result of his intuitions about his language. It is very important to stress right at the beginning that the question under discussion is not whether the speaker *has* the linguistic intuitions (since she obviously does) or *how* intuitions are defined, or *how reliable* they are, but the question under discussion is the following: Where do language intuitions come from? The discussion is primarily and exclusively over *the source of linguistic intuitions*.

There are two main opposing views on the topic about the source of linguistic intuitions, i.e., there is a crucial disagreement what the source of linguistic intuitions is. *Competentialists* (i.e., Miščević) believe and defend the view that linguistic competence is their source. *Ordinarists* (i.e. Devitt and myself) believe and defend the view that linguistic intuitions are not derived from linguistic competence. Intuitions do not flow directly from competence (says the ordinarist) but they are "immediate and fairly unreflective empirical central-processor responses to linguistic phenomena" (Devitt 2006a: 120). The intuitions are mostly the product of experiences of the linguistic world. They are like "observation" judgments.

The issue can be put as follows: a) Linguistic competence is the competence to process language; b) Does linguistic competence also come up with judgments *about* language? Miščević agrees that we should distinguish these two levels: a) dealing with the proposed sentence, at the object-level, and b) reaching the verdict about its grammaticality, at the meta-level.

I was prompted by these words of Miščević and in what follows I will first comment on Miščević's answers to me that I fully disagree with but my main objective is to concentrate on the distinction that Miščević himself makes clear: object-level and meta-level. *The main aim is to make some steps forward since our discussion about linguistic intuitions has come, as I stressed, to a certain not very stimulating stand-still.*

In distinguishing the two levels Miščević goes on to comment on the litmus analogy from Miščević (2014). I have to include a longer quote for the present reader:

Miščević says:

> Let us assume that the person successfully produces the sentence. Is not that success like the information about the color of litmus? (2014: 148)

My answer was:

> Yes, sure it is—but this is part of the production—either of litmus or of competence—which does not carry with it any judgement! (2014: 148)

Miščević's answer to the above is the following:

> But this is all I need at the object-level. Namely, the 'parser in the competence' analyses the sentence and ends up with some mental equivalent of the tree diagram. And this is the main job to be done. If competence is doing this, then it plays the main role. And if Dunja admits this, how can she be an ordinarist, rather than a competentialist? If Devitt agrees with this, how can he be an ordinarist? (2014: 148)

The above is *a very curious* conclusion to say the least since there is NO disagreement between us of what language parser does! We have gone through this point many many times.[2] All that this shows is that competence plays the main role in production and comprehension of sentences. But there is no YES or NO answer in the very production or comprehension — competence does not come out with the judgement about its own doings. On the contrary, whether competence comes out with YES or NO answer is the bone of disagreement and the very difference between an ordinarist and a competentinalist!

Miščević continues:

> Alternatively, further research might show that competence does also issue a verdict, and I hope this is what will happen. My reason for thinking this is that I think competence participates in other linguistic tasks, in particular in immediate linguistic understanding. (2014: 149)

This is even more curious comment since contrary to the previous claim that competence does issue a verdict (YES or NO answer) here Miščević is expressing a possible hope/wish put in future research that might prove that the competence *will* issue a verdict. Since Miščević's main claim is that competence *does* issue a verdict how is the future research going to help?

Miščević continues:

> Let me again state my view. First, in the context of explicit recognition of grammaticality, in which the work of the language module starts by simulating the production of the sentence, and continues by the parsing process, the analytic work is done by the competence, as Dunja agrees. (2014: 151)

Yes I do, apart from the fact that, once again, the workings of competence are *not* in question. I disagree also with the phrasing. I would not call parsing process as 'recognition of grammaticality' but use a more neutral term, 'the production of language strings'.

Then comes the passage where Miščević says: "Further, the general intelligence, and its equivalent, the CP, does not read parsing trees. Therefore, it cannot reach the first verdict" (2014: 151). But this is exactly what the ordinarist is stressing! What is the route from encapsu-

[2] Miščević (2006, 2009, 2012); Jutronić (2012, 2014); Devitt (2006c, 2013).

lated modular workings of the parser to the central processor? As Devitt stresses the explanation would require a relatively *direct cognitive path from the embodied rules of the language to beliefs about expressions of that language*. What could that path be? *There does not seem to be any direct path from these rules to relevant beliefs*.[3]

Commenting of Levelt's suspicion of the role of competence in issuing judgements Nenad says:

> Isn't the performance, i.e. the behaviour, of a pianist often a clear window, almost a clairvoyant window, on her musical competence (and the same for cooking, acting and scoring: Ronaldo's scoring is a clear window on his sporting competence)? (2014: 153)

Surely! But the respective competences are not in question and furthermore they are not the subject of discussion. In other words, Ronaldo's competence is not in the subject of dispute. The question is his (our) judgement about his competence. A different matter.

At this point it became clear to me that there is some possible confusion or misunderstanding about the distinction between the object-level and meta-level and maybe that some progress can be made in our discussion if we see what is really going on here. Miščević prompted me in that direction with his own observation about the analogy of competence with police spokesperson.

Miščević says:

> Let me stay for a second with the police analogy. The spokesperson comments and says: 'Joe was killed by Thomas Mair'. That's an object-level statement. 'And we are quite sure this is how things happened', she might add, either spontaneously, or in response to a question. There, she has passed to the meta-level, and there is nothing unusual about it (2014: 151).

Miščević in criticizing my approach/belief that the main hero for the ordinarist is the central processing unit and that CP has access to the resulting output of a particular competence, so it does some reflection about the output, i.e., about the data provided by the competence says the following:

> There are two assumptions behind Dunja's line of arguing. The first is that there should be a sharp separation between the analysis and the verdict, the object-level and the meta-level. (2014: 150)

And Miščević thinks that "the report of the (alleged) fact and the report about the report's epistemic standing (that 'we' are sure about it) go naturally together" (2014: 151).

My question is: Do they go together? And what does '*naturally go together'* really mean? It seems that the competentinalist (Miščević) wants to minimize the distinction between the object-level and meta-level.

[3] See the section on the route question.

3. *Object-level vs. meta-level*

In what follows I will try to show that competentinalist actually conflates object and meta levels and I want to show the final consequences of such a move. On the other hand, the ordinarist definitely believes that there should be a separation between the object-level and meta-level. *Intuitions are meta-level phenomena and they do not naturally flow from object-level.* Here are some supporting quotations.

Maynes and Gross say:

> But the capacity for linguistic intuitions is a further, indeed dissociable, capacity that goes beyond the capacity for language production and comprehension. It is one thing to parse or understand a sentence; it is another to form an intuition that the sentence is acceptable. (2013: 717)

The psychologist and neuroscientist Barbara Luka says:

> Processes of evaluation are not an automatic result of comprehension or parsing, but rather require attention. The ability to provide linguistic judgments is also dependent on metacognitive and analytic reasoning capacities of individual speakers (2005: 488).

In discussing speaker's introspective experiences Luka wisely warns us that many questions stay unanswered regarding the interaction betweeen linguistic knowledge and the processes required to perform and report a linguistic judgement. For example, what is the nature of the interaction between sentence comprehension and sentence evaluation? Is evaluation of grammaticality an automatic result of parsing? Devitt stresses that the intuitions are declarative knowledge, the understanding, procedural and that these are very different kinds of knowledge here as elsewhere in our psychology. The behaviors are linguistic *performances* whereas intuitions are *judgments about* language. Two levels have been distinguished for sure and the ordinarist provides the explantion for this.

What about competentionalists?

Miščević's previous answer that "the report of the (alleged) fact and the report about the report's epistemic standing (that 'we' are sure about it) go *naturally together*" indicates that the two levels are brought very close together. Even more, it seems that the object-level and meta-level are almost conflated. It is a conflation of behavior/performance and intuitions about behavior/performance. Miščević says that there is 'nothing unusual' about this. But how can one justify and explain this conflation? According to the mentalist conception, linguistic intuitions are the product of a modularized language faculty that alone delivers the relevant information to mechanisms responsible for judgment. But how does it do it? The answer seems to be 'naturally' but that is not a good or satisfactory answer, especially if we take seriously what philosophers and most importantly psychologists are warning us about[4],

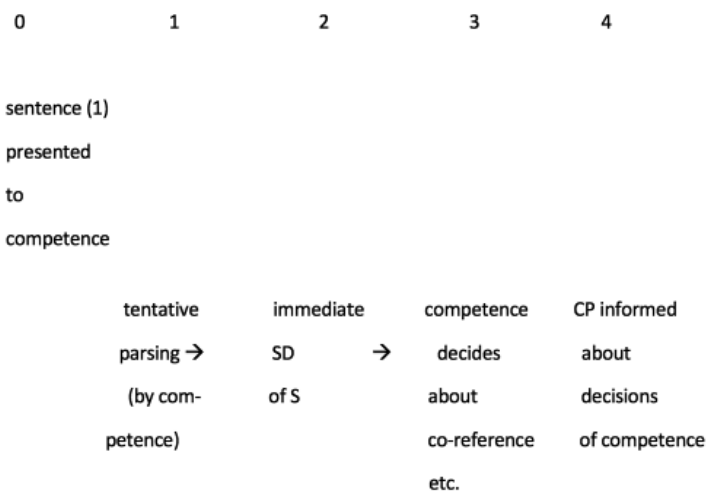
[4] See the above quotes.

i. e. that that not much is known about the mechanisms implicated specifically in the formation of linguistics intuitions.[5]

Full discussion would involve the burning question about the route from the underlying competence to the central processor and that would be a different paper, much more complex one. Nevertheless, let me comment briefly on Miščević' s and Rey's attempts at providing the answer. Let us call it 'The Route Question'.

4. *The Route Question*

Miščević thinks that object-level and meta-level levels come together. In his chart the (core) intuition/ judgment is already made by competence (stage 3 on the flow chart below).[6] Yes or No answer (presumably from stage 3) is simply passed to the central processor (stage 4 on the flow chart). If the central processor does not add anything then one can say that it is in a sense bypassed completely, i.e., does not play any role in forming (narrow) linguistic intuitions. Object-level (parsing, competence decision) and meta-level levels (decision simply passed to CP) are not distinguished.



[5] On the vexing point that the intuitions might be seemings Maynes and Gross say: "The distinction between judgments and appearances is arguably of limited significance, however, to linguistic practice. First, in an overwhelming number of actual cases of interest to linguists, subjects will judge things to be just as they experience them as being. *Second, in any event, if we are to make use of seemings in theorizing, they must be reported.* Thus, a judgment is required after all, albeit to the effect that things so seem to one. (Glüer (2009) argues that perceptual experiences just are judgments about how things perceptually seem.)" (2013: 716, italics mine). Devitt (2010) also argues that there are no such seemings.

[6] The chart from Miščević (2104: 149).

Georges Rey in response to Devitt's claim that hearing an utterance in a certain way is one thing, judging that it has certain properties, admits that it is true that hearing is one thing and judging quite another: nevertheless, Rey comments, sometimes our best judgments are based on what you hear. Again near conflation of object (hearing) and meta (judging what is heard) level!

Mišćević, and especially Rey, put a lot of stress on competence delivering structural descriptions (SD, stage 2 on the given chart above) on which (supposedly) judgements are made by the central processor.

The question that Devitt asks is the following: Suppose that the language system did deliver a partial SD (structural descriptions) to the central processor, how would the SD's information 'fairly directly cause' the intuitions that are the concern of VoC? The question stands unanswered. Rey's answer to Devitt's question is that this is 'trivial'. He claims that going along with the modularity story, there's no reason that attention can't be drawn to the informational outputs of a module, say, by it simply getting highlighted, i.e., computationally enhanced, or sent to a special attention address or something.[7] This again is no good answer, not to mention no explanation at all, since *nobody* has ever claimed that the output of the competence comes highlighted or with some kind of enhanced information apart from simple deliverance of SD's. In other words, nobody has claimed that SD's come with a sign saying "ungrammatical" or "grammatical" or Miščević's YES or NO answers. What is important for our discussion is that it is evident that the object-level and meta-level are conflated.

On competentialist's view SD's are somehow (naturally, not unusually) miraculously 'mainlined' in(to) the central processor. On the ordinarist view there is no mystery. The object-level and meta-level are kept distinct and here is a very short story. I quote Devitt: "I argue that intuitive judgments about language, like intuitive judgments in general, are empirical theory-laden central-processor responses to phenomena, differing from many other such responses only in being fairly immediate and unreflective, based on little if any conscious reasoning" (2006a: 103). On this view, there is no unexplainable cognitive penetration of the central system or SD being 'computationally enhanced, or sent to a special attention address or something'. The object level (competence and the data that it provides) are kept separate from the content of the judgment, i.e. intuitions, that are in the central processor (the meta level). The central processor is the home of inferences and judgements, including intuitive ones.

On the final note, it is instructive to quote an interesting and revealing part from the correspondence between Rey and Devitt on the very question that is being discussed.[8] Rey says that Devitt allows that

[7] The best and the most relevant place for this dispute is Devitt (2013) and Rey (2013) but there are also some email exchanges among them available at request.

[8] Email correspondence but see also Rey (2013) and Devitt (2013).

syntactic properties can cause intuitive judgments as responses and that Devitt just thinks this is mediated by the person's central appreciation of some bit of the theory of grammar, folk or otherwise. But the Chomskyan (competentionalist) argues there is a more direct route: we are *so wired* that the outputs of the language module more directly cause some of our intuitive responses, without the mediation of a *central* theory.

Thus, not only that the object-level and meta-level have been brought closely together (which was Nenad's suggestion) but they have been conflated altogether and moreover finally reduced to an innate instinct (we are so wired)! This seems to me to be a desperate move, not to mention that this is no explanation at all.

5. *Conclusion*

1. I have argued that most of Miščević's (2014) arguments to my objections (Jutronić 2104) are not well founded.

2. In distinguishing and stressing the difference between the object-level from the meta-level I have tried to show that there is a failure from the competentionalist to keep our access to linguistic data provided by competence (object level) sharply distinct from our alleged access to linguistic information provided by competence (Devitt 2010b). There is access to data provided by competence on anyone's view (object-level)! But the ordinarist rejects the view that we have access to linguistic information (meta-level) since this information does not reside in competence.

3. Competentionalist claims that linguistic intuitions are largely supplied by linguistic competence with the penetration from the central processor. Ordinarist claims that the intuitions are supplied solely by the central processor and there is no question of the competence being cognitively penetrated. The distinction between source of data and information is crucial to ordinarist's discussion. The data for judgments (object-level) are not judgments (meta-level). Hearing an utterance in a certain way is one thing, judging that it has certain properties, another. One might comment that there might not be an abyss, or even sharp distinction between the two but one cannot deny that there is at least a 'thin red line' between the two, and that line makes all the difference in finding out who is right in this debate.

References

Devitt, M. 2006a. *Ignorance of Language*, Oxford: Clarendon Press.

Devitt, M. 2006b. "Intuitions in Linguistics." *British Journal for the Philosophy of Science* 57: 481–513.

Devitt, M. 2006c. "Defending *Ignorance of Language*: Responses to the Dubrovnik papers." *Croatian Journal of Philosophy* 6: 571–606.

Devitt, M. 2010. "What 'Intuitions' are Linguistic Evidence?" *Erkenntnis* 73: 251–64.

Devitt, M. 2013. "Linguistic Intuitions are not 'the Voice of Competence'." In Haug 2013: 268–93.

Devitt, M. 2014. "Linguistic Intuitions: in defense of 'Ordinarism'." *European Journal of Analytic Philosophy* 10 (2): 7–20.

Devitt, M. Forthcoming. "Linguistic Intuitions: A response to Gross and Rey." In S. Schindler (ed.). *Linguistic Intuitions*. Oxford: Oxford University Press.

Glüer, K. 2009. "In Defense of a Doxastic Account of Experience." *Mind & Language* 24: 297–327.

Haug, M. (ed.). 2013. *Philosophical Methodology: The Armchair or the Laboratory?* London: Routledge.

Jutronić, D. 2012. "Je li jezična sposobnost izvor jezičnih intuicija (Is languge competence the source of linguistic intuitions)." In S. Prijić-Samaržija and P. Bojanić (eds.). *Nenad Miščević – Sva lica filozofije (Nenad Miščević – all faces of philosophy*). Beograd: Institut za filozofiju i društvenu teoriju: 129–143.

Jutronić, D. 2014. "Which Are the Data That Competence Provides for Linguistic Intuitions?" *European Journal of Analytic Philosophy* 10 (2): 119–143.

Luka, B. 2005. "A Cognitively Plausible Model of Linguistic Intuitions." S. Mufwene, E. Francis, and R. Wheeler (eds.). *Polymorphous Linguistics: Jim McCawley's Legacy*. Cambridge: MIT Press: 479–502.

Maynes, J.and Gross, S. 2013. "Linguistic Intuitions." *Philosophy Compass* 8 (8): 714–730.

Miščević, N. 2006. "Intuitions: The Discrete Voice of Competence." *Croatian Journal of Philosophy* 6: 523–548.

Miščević, N. 2009. "Competent Voices: A Theory of Intuitions." http: oddelki.ff.uni-mb.si/filozofija/en/festschrift.

Miščević, N. 2012. "Odgovor Dunji Jutronić (Answer to Dunja Jutronić)". In S. Prijić-Samaržija and P. Bojanić (eds.). *Nenad Miščević – Sva lica filozofije (Nenad Miščević – all faces of philosophy*). Beograd: Institut za filozofiju i društvenu teoriju: 194–198.

Miščević, Nenad. 2014a. "Reply to Michael Devitt." *European Journal of Analytic Philosophy* 10 (2): 21–30.

Miščević, N. 2014b. "Reply to Dunja Jutronić." *European Journal of Analytic Philosophy* 10 (2): 145–53.

Rey, G. 2006. "Conventions, Intuitions and Linguistic Inexistents: A Reply to Devitt." *Croatian Journal of Philosophy* 6: 549–69.

Rey, G. 2013. "The Possibility of a Naturalistic Cartesianism Regarding Intuitions and Introspection". In Haug 2013: 243–67.

Schindler, S. (ed.). Forthcoming. *Linguistic Intuitions*. Oxford: Oxford University Press.

Croatian Journal of Philosophy
Vol. XVIII, No. 53, 2018

Devitt's Promiscuous Essentialism

ZDENKA BRZOVIĆ
University of Rijeka, Rijeka, Croatia

In this paper I examine Michael Devitt's version of essentialism, a view that stirred a lot of debate amongst philosophers of biology by going against the mainstream view of "death of essentialism" in evolutionary biology. So far, much more attention was directed to refuting Devitt's view then to analyzing what his essentialism consists in. I go through the main tenets of the essentialist view, examine the relation between Devitt's view and the so-called traditional essentialism, and the cluster approaches to natural kinds. I conclude that Devitt holds a very flexible variety of pluralistic essentialism, that I term promiscuous essentialism. The benefit of holding such a view is that it can encompass a wide range of categories, but its downside is that knowing the essence of a kind can be minimally explanatory. For this reason, the criterion for privileging certain kinds cannot follow from identifying their essence, which was originally one of the main motivations for holding an essentialist view.

Keywords: Natural kinds, species, essentialism, pluralism, philosophy of biology, cluster kinds.

1. *Introduction*

Michael Devitt in his thought-provoking paper "Resurrecting Biological Essentialism" argues that species and Linnean taxa more generally have essences that are, at least partly, constituted by intrinsic properties (Devitt 2008a). Many philosophers of biology reacted to his view by defending the near consensus view, according to which, at least when it comes to species, there is no place for essentialism in biology (Barker 2010; Ereshefsky 2010; Lewens 2012). Most of the reactions did not reflect upon Devitt's version of essentialism as being somehow specific or different from the traditional variety. In this regard, Marc Ereshefsky's (2010) discussion presents an exemption. He identifies Devitt's view as belonging to "new biological essentialism", but does not discuss what traditional essentialism consists in, and how the new biological essentialism differs from it. The aim of this paper is to offer an analysis

of Devitt's essentialism and to situate it in the taxonomy of possible essentialist positions and examine its relation to cluster kinds view of natural kinds.

Essentialism is normally taken to have three core tenets: (1) all and only kind members possess a common essence; (2) essence causes traits typically associated with kind members; and (3) identifying an essence helps us to explain and predict traits associated with kind members (Ereshefsky 2017). However, if we examine further the platitudes associated with the essentialist view we will find two additional claims: (4) that essences ought to be intrinsic properties of kind members, as opposed to being extrinsic or relational ones, and (5) natural kind monism, i.e. the claim that there is one correct way of dividing the world into natural kinds[1] (see, for instance, Wilkerson 1993). I will take these five tenets to make up the traditional essentialist view.

In the next section, I will present some reasons for endorsing (4) and (5) in addition to the three basic essentialist tenets. After that, I will distinguish essentialist approaches from cluster approaches to natural kinds and argue that, regardless of some authors like Ereshefsky (2010) and perhaps even Devitt (2008a) who include them among essentialist views, they cannot be considered as such because they defy tenet (1) of essentialist positions. Thus, I will argue that if Devitt's view is construed as essentialist then it cannot be considered as belonging to cluster approaches to natural kinds. Devitt does not elaborate much what is the specific brand of essentialism that he endorses. However, he does mention that Kripke's (1980), Putnam's (1975) and Wiggins's (1980) type of essentialism represents a received view in philosophy, with the exception of philosophy of biology. Thus, I use this as evidence for interpreting what type of essentialism Devitt wishes to resurrect. I will use Putnam here as the relevant source and argue that both Devitt and Putnam endorse a pluralist version of essentialism, thereby denying tenet (5) of the traditional essentialist view. The I will argue that while for Putnam it is not entirely clear how broad his pluralist view is, it appears that Devitt's essentialism is very encompassing. In fact, his essentialism seems so broad that it resembles John Dupré's promiscuous realism, which is why I term his view promiscuous essentialism. Finally, I will indicate some problematic consequences of such a view. Namely, Devitt's essentialism seems to be so relaxed that the essence of a kind could be a property that plays a minimally explanatory role.

[1] Natural kind monism is rarely explicitly stated in this form (although, see above quoted Willkerson) but can be found in accounts that emphasize that natural kinds should form a hierarchy; the view that there should be no cross-cutting kinds, and that if two kinds overlap one should be a subkind of the other (see, for instance, Ellis 2001). While in priciple, this hierarchy thesis is compatible with pluralism in case we have different hierarchical systems of natural kinds, normally it is assumed that there is one correct hierarchy, and that there should be no cross-cutting classifications whatsoever. This assumption will be taken for granted in the rest of the paper, but this issue is discussed further in section 2.

2. *Traditional essentialism*

Essentialism requires that all and only members of a natural kind have a certain property—essence—which is, in turn, responsible for the traits characteristic for that kind. To illustrate, take Putnam's famous example that the essence of water is H_2O. The molecular structure of water is what makes all instances of water members of the same kind. Moreover, this molecular structure is causally responsible for other, characteristic properties of water, such as being colorless, odorless and transparent. Thus, discovering the essence of a kind allows us to explain and predict properties of members of a kind. This sums up the three main tenets of essentialist views. Two further claims that have been typically associated with essentialism are (4) that essences ought to be intrinsic properties of kind members; and (5) natural kind monism.

With regards to (4), the paradigmatic examples of essentialist kinds are taken to have intrinsic properties. That is, properties associated with a natural kinds will be possessed by members of that kind independently of their relation to other things (Ellis 2001; Wilkerson 1988). Take for instance chemical elements. It is taken that the essence of chemical elements is their atomic number, or, to be more specific, properties of the nucleus of atoms belonging to an element. Gold, for example, has the atomic number 79, which means that it has 79 protons in the nucleus, and the same number of electrons in the extranuclear region of the atom. Since the chemical characteristics of gold (and other elements) are closely related to the number and arrangement of electrons in their atoms, it is taken that elements are entirely distinguishable from each other by their atomic numbers. The basic idea is that our scientific investigations of the world will lead to classifications that reflect the discontinuities and boundaries between natural objects and processes. Thus, upon encountering an entity or process, and after examining its intrinsic properties, we should be able to decide to which kind(s) it belongs.

One reason for excluding extrinsic or relational properties as essences of natural kinds is that very heterogeneous sets of entities can enter into the same relations and possess common extrinsic properties. Functional kinds, for instance, are defined by invoking relational or extrinsic properties; entities are classified into a kind because of something they (can) do, a function they can serve, and not because of some intrinsic similarities. For instance, a watch is a functional kind because it is defined by its function to measure time. A watch can be digital, analog, mechanical, electronic, and so on and so forth. Objects made out of different materials are watches if they serve a function in relation to measuring time. Similarly, money is a functional kind because it is individuated by its role in human economy and not by intrinsic properties of objects that can play this role (paper, metal, digital currency, etc.). For this reason many functional kinds are taken to be non-reducible to natural kinds (Fodor 1974).

A variation of essentialism in chemistry is microstructuralism, the view that microstructural properties make up the essence of chemical kinds. In the case of chemical elements this can be the nuclear structure, but it remains to be specified what exactly this structure consists in, or, how to pinpoint what constitutes microstructural similarity. If we wish to focus on nuclear charge, for instance, then classification according to atomic number (i.e. the number of protons in the nucleus) will be relevant. On the other hand, if we focus on nuclear mass, then the number of neutrons will be relevant as well, and we will base a classification on more fine-grained isotopes of elements. Or, we can classify substances by focusing on the patterns of radioactive decay and reach categories such as radionuclides which cross-cuts classification into chemical elements. The possibility of different microstructural features being essences of different natural kinds opens up the question whether there is one correct division into kinds according to microstructural essences. This brings us to tenet (5) associated with essentialist views, namely that there is one correct way of dividing the world into natural kinds according to their essences, i.e. natural kind monism.

On the monistic view no cross-cutting classifications can be considered natural kinds. In case there is an overlap between natural kinds categories, one of them ought to be a subkind of the other. That is, there should be no cross-cutting categories. For instance, the category of ubiquitous organisms refers to organisms that can tolerate a wide range of environmental conditions and it comprises diverse organisms such as bacteria and fungi. However, it cross-cuts standard biological taxonomy according to which bacteria and fungi belong to different biological domains, bacteria comprising their own domain, and fungi belonging to the domain of eukaryotes. In such cases, the natural kind monist ought to conclude that at most one of the cross-cutting classifications represents a natural kind. The only overlap can happen between categories belonging to the same hierarchy where one is a subcategory of the other. For example, humans belong to the category Homo sapiens, but also to the categories Mammal and Vertebrate, because Homo sapiens is a species belonging to a class of Mammals, and they both belong to subphylum Vertebrates. This view is called a *hierarchy thesis* regarding natural kinds and it is often suggested as one of the criteria that natural classifications ought to fulfill (Bird and Tobin 2017). Monists who do not endorse the hierarchy thesis ought to claim that we can find natural kinds only on the lowest level of classifications, and at that level there should be no overlap between kinds.

Essentialism is typically perceived as a monistic approach to natural kinds. This can be illustrated by Hilary Putnam's view on natural kinds. His position can be qualified as essentialist in the light of the fact that he defines natural kinds as consisting of individuals that bear sameness relation to the specified paradigmatic exemplar of the kind. However, he also adds that the sameness relations we use to identify

natural kinds are interest relative. More concretely, he claims that, for instance, one thing holds the "same liquid" relation to something else if the two agree in important physical properties, where importance is an interest relative notion (Putnam 1975a). Thus, in our everyday contexts it is correct to say that water is H_2O because for everyday contexts the important properties of water are captured by the compound H_2O. But when we are doing chemistry then "the same liquid as" also refers to other molecular structure because water, beside H_2O, consists of D_2O, D_4O_2, D_6O_3, etc.

Ian Hacking (2015) argues that Putnam's insistence on interest-relativity of natural kinds commits him to the view that kinds lack essences. Hacking's interpretation of Putnam is plausible if we assume that essentialist natural kinds ought to be fixed categories that do not depend or change with our interests, and once we identify the kind's essence we have clearly established the demarcating lines of that kind. The assumption is that neither demarcation nor essence is something that can change with our interests. I introduced Putnam's view here exactly because he is usually considered as a typical proponent of essentialism, which, by many interpretations, goes hand in hand with natural kind monism. Thus, Wilkerson (1993: 4, 5), when discussing and defending Putnam's position against Dupre's criticism says that it and other "doctrines of natural kinds" hold that "although there are many similarities and differences between things, one set of similarities is privileged, because they are the real essences which determine natural kinds." Commitment to monism is often not stated explicitly in the traditional essentialist views, but can be inferred from the fact that they typically endorse the hierarchy thesis (see for instance, Ellis 2002).

On this reading essences ought to be somehow special or privileged properties, and once we identify them, we know what the uniquely appropriate way of dividing the world into natural kind categories is. On the other hand, approaches that put emphasis on the importance of the fact that categories ought to serve our interests are pluralistic, because in different contexts and disciplines our interests can vary and with them the categories we deem natural. It seems highly unlikely that the view that puts focus on interest-relativity of groupings will arrive at a monistic division of natural kinds.

In section 4 I will discuss Devitt's view as a type of pluralistic or promiscuous essentialism, because he appears to allow for a wide range of interests to play a role in demarcating natural kinds. However, before engaging this issue I will argue that there is an important distinction between essentialist and cluster approaches to natural kinds and that Devitt's view can be interpreted as essentialist with this distinction in mind.

3. *Essentialism vs. cluster kinds*

The main argument against the traditional essentialism in philosophy of biology relied on the strictness of the requirement that all and only members of a kind should share an essence. Given the evolutionary history of different organisms, it is unlikely that we will find such a property that is unique to kind members and that will provide grounds for biological classification into species. In fact, according to this argument, even if there were such a property it would likely disappear from a population of organisms given the workings of evolutionary forces such as mutation, recombination or random drift (see, for example, Ereshefsky 2017, Okasha 2002).

Devitt (2008a), against this consensus in the philosophy of biology, claims to be resurrecting species essentialism. However, it is not clear what kind of essentialism he seems to be resurrecting. The aforementioned anti-essentialist consensus argues against species being defined by intrinsic essences and claims that species are defined by relational properties (some even argue for extrinsic or relational essentialism (see, for instance, Okasha 2002)). Take one of the more popular species concepts, for example, the Biological Species Concept (BSC) defines species as members of populations that can potentially interbreed. In this case, the potential to interbreed with other species members is a relational property which specifies species membership, and not some intrinsic property that species members share. Devitt, on the other hand, claims that species members, and Linnean taxa in general, share at least some intrinsic properties and that species concepts are not entirely relational. This, by itself does not seem sufficient to establish an essentialist view. Anti-essentialists do not have to deny that species members can share some important intrinsic properties. Nonetheless, they put emphasis on variation between individuals given the operations of evolutionary processes, and, thus, claim that species are distinguished by "clusters of covarying [chromosomal and genetic] traits, not by shared essences" (Okasha 2002: 197). Devitt seems to endorse this view but gives it a different spin. He quotes exactly this statement by Okasha and adds that the clusters in question are exactly the essences he is talking about. This brings us to the interesting question about the relationship between essentialism and cluster approaches to natural kinds.

There are at least two possible ways to interpret this relationship, depending on how strictly we define essentialism. We can identify essences with necessary and sufficient conditions for kind membership (Magnus 2012), which I take it corresponds to tenet (1) of essentialist views: all and only kind members possess a common essence. According to P. D. Magnus, and I tend to agree, this is the main feature of essentialist accounts. In addition, Magnus claims that this criterion amounts to the assumption that natural kinds ought to have sharp boundaries,

i.e. be categorically distinct (Magnus 2012: 19). This means that for any individual entity, it must be clear whether it is a member of a certain kind or not.

An argument that has been worked out for the sharpness or categorical distinctness of essentialist natural kinds relies on the intuitions that natural kinds ought to pick out real features of the world. Accordingly, Brian Ellis (2001: 19, 20), for instance, argues that if natural kinds were continuous, and thus, not categorically distinct, then it would be up to us where to draw the line where one kind ends and another begins. This would make the delimitation of natural kinds a matter of convention, in opposition to essentialist claim that they are determined by *real* features of the world.

The cluster views were introduced exactly with the intention of accommodating the fact that many natural kinds are not categorically distinct, and their aim was to work out a more encompassing account that would capture many actual scientific categories. Take the example of species, members of such kinds tend to share many common properties, but no property is unique to them. For instance, black stripes are characteristic of tigers. However, there are also tigers that do not have them. Accordingly, a specific property or a well-defined set of properties characteristic for a cluster kind is not a necessary condition for an entity to belong to that kind and thus natural kind boundaries can be vague. For instance, dogs and wolves have many similarities and they can have viable offspring. But, given their habitats, social structure and many other typical features it is not clear that it is useful to consider them the same species or natural kind.

There is however, another interpretation of the relation between essentialism and cluster views that does not require essences to be unique, and natural kinds to be categorically distinct. By essence we can understand only that there are some facts about the world corresponding to the unity of the kind, as one interpretation would suggest (Magnus 2012). On this reading, anyone who believes that there are natural kinds at all, thinks that they have this type of essences, and we end up with a position that is even more encompassing than cluster kinds view. This takes away significance from essentialist views and merely equates them with positions holding that natural kinds ought to possess some unity that corresponds to some facts about the world. If we interpret Devitt as holding this view, then not much resurrecting has been done on his part since cluster kinds accounts are popular in philosophy of biology (Boyd 1999; Griffiths 1999; Robert A. Wilson, Barker, and Brigandt 2007).

We need to address the possibility, however, that Devitt has a different view of the consensus in philosophy of biology, and that he thinks cluster accounts of natural kinds are considered outdated or simply wrong. He talks about clustering views, i.e. the HPC view, as holding, as he does, that species have at least partly intrinsic essences.

But then he quotes Paul Griffiths (1999) as arguing that species have purely historical essences which is then interpreted by Devitt as being incompatible with the HPC theory (Devitt 2008a). This is interesting since Griffiths is standardly taken as a proponent of the HPC theory (see, for example, Ereshefsky 2017). Explaining Griffiths' view should help to make more precise the difference between clustering and essentialist accounts.

Griffiths (1999) argues that causal homeostatic mechanisms play the same role as essence plays in traditional essentialist accounts—that is, on this new understanding, essence refers to the states of affairs that license induction and explanation within a theoretical category. Thus, for some types of scientific categories, the role of essence can be played by entirely extrinsic or relational properties. Griffiths thinks that this is compatible with the HPC theory and offers the example of money. There can be indefinitely many physical instances of money (or even non-physical ones) but the essence, i.e. the reason why all these instances are considered money is that they are recognized and evaluated as such by the consumers, that is, by this relational property.

When it comes to species, however, he claims that in addition to relational or historical essences, members need to share some intrinsic properties as well. This is clear when he talks about causal mechanisms, such as developmental ones, responsible for, for example, prey detection in certain species of birds (Griffiths 1999). Perhaps the best way to put it is to say that the extrinsic essence is responsible for the fact that species members share intrinsic properties as well. Griffiths explains why kinds whose only essential properties are historical should be subjects of lawlike, counterfactual-supporting generalizations about morphological and physiological properties; because the principle of heredity acts as a kind of inertial force until some adaptive force acts to change that form (see his explanations of *phylogenetic inertia* in support of this view (1999: 220)). This is consistent with using phylogenetic relations as a species essence, for instance, Phylogenetic Species Concept (PSC) identifies species as sets of organisms sharing a common ancestor. Another way that extrinsic essence can be responsible for sharing intrinsic properties is by enabling the exchange and sharing of genetic material causing many of the shared traits (BSC species concept limits species members to those that are able to exchange genetic material trough reproduction). I take it that the other proponents of the so-called relational essentialism about species hold similar views. This is consistent with Mayr's (1961) distinction between ultimate and proximate causation and the corresponding two sorts of explanation, which Devitt cites approvingly (Devitt 2008a: 353).

What is the difference, then, between relational essentialists[2] and Devitt, if they also hold that species members need to have some in-

[2] Griffiths (1999), Okasha (2002) and LaPorte (2004) are taken as proponents of relational essentialism.

trinsic properties in common, in addition to relational ones? I take it that proponents of such accounts do not call the intrinsic properties in question species' essences, exactly because on traditional essentialist accounts, essences refer to necessary and sufficient properties for species membership. While phylogenetic relationships can be taken as sufficient and necessary conditions for belonging to a certain species on certain species concepts, it is not so easy (or it is perhaps impossible) to specify the necessary and sufficient intrinsic properties that species members ought to share because of variation between them. In other words, while it is, at least on some species concepts, clear to what species an organism belongs to, just from knowing a certain (important) relational property,[3] this is not clear just from examining organism's intrinsic properties. Thus, I take it that relational essentialism would agree that species members do share common clusters of properties (and they share them because of the relational essence), but we cannot specify any set of such properties that will be unique only to members of one species. If we wish to take Devitt's view as arguing for something stronger than this, then he must be committed to the claim that we can delimit a set of necessary and sufficient intrinsic properties for species members. Otherwise, it is not clear how intrinsic his biological essentialism is.

In the next section I attempt to characterize in more detail Devitt's version of essentialism on the assumption that he does endorse intrinsic essentialism.[4]

4. *Devitt's promiscuous essentialism*

As was illustrated in section 2, Putnam's essentialism, which Devitt takes as paradigmatic essentialist view in contemporary philosophical debates and seems to rely upon in his own essentialist view, is pluralistic, in opposition to the traditional essentialism. Pluralists about natural kinds hold that we can arrive at many different, cross-cutting classifications of the entities (and/or processes) in the world. Depending on what we are interested in, we will arrive at different classificatory systems. For instance, if we are interested in patterns of radioactive decay we will arrive at a classification that cross-cuts the standard

[3] Be it that the organism shares a common ancestor with other members of that species or that it can interbreed with them, the reader can fill out here her favorite relational species concept if the concept allows for clear cut distinction between different species.

[4] Perhaps his invoking of clusters has a different purpose; to point out that the essence of a species need not correspond to a crude idea that there is one gene that makes a tiger a member of the tiger species, for example. Thus, when he talks about clusters or patterns of properties and that the intrinsic essence does not need to be neat and tidy, he might have in mind the fact that essences can be very complex and comprise various properties that come together and make up a species. This is compatible with the view that essences are unique properties that all and only kind members share.

chemical classification into chemical elements, which is suitable for many other interests such as explaining material transformations. There is nothing inconsistent in claiming that we can have essentialist classifications that allow cross-cutting categories, if the essences in question are uniquely shared by members of the kind. For instance, category vitamin A is defined by its biological properties in a vitamin-deficient organism, but it consists of at least six vitamer chemicals that differ in their chemical structure, so the category cross-cuts standard chemical classifications. In this case, the specific biological activity of the vitamins can be considered as their essence, even though it does not correspond to the microstructural essence of the compounds comprising the category.

Devitt allows for a vast range of categories to be considered essentialist, and not just a limited set of basic physical, chemical or biological classifications. He shares the basic pluralist intuition that, depending on our interests, we might carve out the world differently and that different properties will make up essences of those kinds. This can be seen from the example he offers to support his claim that essences can be partly intrinsic and partly relational, or entirely relational. The essence of being a pencil is partly determined by its relation to human intentions and partly by its physical properties, the essence of being Australian is entirely relational or extrinsic.

That there is a certain amount of terminological misunderstanding in this whole debate can be seen from the fact that Devitt's essentialism, as interpreted here, is stating something very similar to Dupré's promiscuous realism. Dupré, however argues against essentialism. His promiscuous realism is a claim that there are many sameness relations that can determine kind membership, and which ones will be taken as relevant will depend on our interests in various circumstances. He puts both scientific and every day or folk categories on equal footing in the sense that they all can be considered natural kinds. Devitt's examples of essentialist kinds such as being Australian, or pencil, goes in the similar direction. Only, on my interpretation, his view should not allow vague boundaries between categories, while Dupré's does. While this fulfils the minimal essentialist requirements, it downsizes the role of essences to a considerable degree. Essence was originally assumed to be important in grounding the explanatory success of natural kind categories. This is what Devitt relies on in his criticism of the anti-essentialist consensus regarding species; he claims that being a member of a biological taxon ought to be explanatory.

If we take as core of essentialism the three main tenets (1)–(3), even these mentioned categories can fulfil them. Accordingly, if the categories 'pencil' and 'being Australian' are defined strictly enough, this criterion will yield clearly demarcated categories where all and only members of a kind share an essential property. For example, if we define pencils as instruments for writing that consist of a solid pigment core

inside a protective casting, and if the definition is strict enough, we will have a clearly delineated category of writing instruments where all members share an essential property. Also, that essence will be responsible for (some) traits associated with that category members (tenet 2), and by identifying it we can explain and predict properties of category members (tenet 3).

With regards to biological classifications Devitt says he defends the doctrine of intrinsic biological essentialism, according to which Linnean taxa have, at least partly, underlying intrinsic properties. Linnean taxa are not just species, which are standardly taken as candidates for natural kinds, they also include all categories in the Linnean hierarchy such as kingdoms, phyla, classes, orders, families and genera. Take the example of the kingdom Archaeplastidans, they are characterized by having plastids—chloroplasts that carry out photosynthesis and are derived from captured cyanobacteria. If this is a feature of all the organisms classified in this kingdom, we can take it to be the essence of that category.

It is questionable whether there is such a feature for all the taxonomic ranks, but I take it that an essentialist about taxonomic ranks must argue that there is such a feature. A potential problem with such an approach is that, while it is compatible with tenet (1) of essentialist views, it is questionable to what degree it can be taken to fulfill tenets (2) and (3). Namely, the fact that all Archaeplastidans have plastids is responsible for some of their shared traits, but there is also much divergence in other traits of these organisms. For instance, they can vary from being isolated cells to colonies and multi-celled organisms. Thus, such categories have a very limited predictive and explanatory value.

To use another, more familiar example, how explanatory is the category of vertebrate? It is taken to comprise all species with a backbone, but it includes such diverse organisms as Fire salamander, Saltwater crocodile and House sparrow. Knowing that an organism is a member of this category is minimally explanatory because all it can explain is a few facts about the most general features of its body plan. If we wish to provide an account of natural kinds as explanatory categories, then either the essence should be some very important property that causes many other properties of kind members, or, the essence is not what makes such categories explanatory in the first place.

Even more extreme case is Devitt's own example of 'being Australian' as a relational or extrinsic essentialist category. While this might formally be considered as an essentialist category, the minimal number of characteristics of category members that the essence (however we might specify it) is responsible for, has next to zero explanatory and predictive value, which can make us wonder whether it is justified to call these essentialist categories. While a certain amount of pluralism surely is compatible with essentialism, regardless of this not being the traditional approach to essentialism, this type of promiscuous essen-

tialism seems to go too far, which can make us reconsider it as an essentialist view.

Admittedly, it is not clear whether Devitt himself would consider all the aforementioned essentialist categories as natural kinds, since he thinks that natural kinds need to be *explanatory significant* (Devitt 2008b). Now, he says that explanatory significance comes in degrees, but we might agree that the kind 'being Australian' has next to zero explanatory value and can therefore be excluded from the category of natural kinds. If Devitt wants a way out of this type of promiscuous essentialism, he needs to offer a criterion, based on which we can decide what categories are fulfilling the *explanatoriness* requirement. When he talks about 'carving nature at its joints' Devitt mentions that kinds of entities posited by a scientific theory ought to play a *causally significant role* but does not elaborate further on how to recognize and delineate such causally significant roles.

Traditionally, essences fulfilled that function. Recognizing a kind's essence allows us to establish that kind as genuinely explanatory and the essence in question is what grounds this explanatoriness. If, however, essences are downsized to such a degree that even 'being Australian' can be an essentialist kind, then essence no longer plays the same role. Let us go back to the example of species, how can we tell that *Canis lupus familiaris* plays a causally significant role, or possesses an essence? It appears that Devitt's response is that it must possess it, because it is obviously explanatorily significant, and, consequently, it is the job of working biologists to find specific essences of particular species. This seems like it inverts what the main upshot of essentialism was supposed to be; that finding an essence will provide the grounding for the explanatoriness of natural kind categories. If we start out by stating that certain categories are obviously explanatory and conclude from this that they must possess an essence, then we have no criterion of how to distinguish genuinely explanatory categories from the ones that are not explanatory, other than some common-sense estimation. This strategy can easily lead into the promiscuous variety of essentialism.

5. *Conclusion*

I have analyzed Devitt's version of essentialism and its relation to the traditional essentialism and the cluster accounts of natural kinds. I have argued that his variety of essentialism is either too promiscuous, or, he needs to offer a criterion of what makes natural kinds genuinely explanatory in opposition to any everyday classifications that share some common property. While traditionally in essentialist views, essence was supposed to play that role, Devitt's downsized notion of an essence appears unsuitable for it.

Acknowledgments

This paper was presented at the conference "Devitt's 80th" (University of Maribor, May 9–10 2018). I thank the audience and especially Michael Devitt for their helpful comments on a previous version of this paper. This work is an output of the project CEASCRO (grant HRZZ-IP-2013-11-8071), funded by the Croatian Science Foundation.

References

Barker, M. J. 2010. "Specious Intrinsicalism." *Philosophy of Science* 77 (1): 73–91. https://doi.org/10.1086/650209.

Boyd, R. 1999. "Homeostasis, Species, and Higher Taxa." In R. A. Wilson (ed.). *Species: New Interdisciplinary Essays*. Cambridge: MIT Press: 141–85.

Devitt, M. 2008a. "Resurrecting Biological Essentialism." *Philosophy of Science* 75 (3): 344–82. https://doi.org/10.1086/593566.

_____2008b. "Biological Realisms." In H. Dyke (ed.). *From Truth to Reality: New Essays in Logic and Metaphysics*. London: Routledge.

Ellis, B. 2001. *Scientific Essentialism*. Cambridge: Cambridge University Press.

_____2002. *The Philosophy of Nature: A Guide to the New Essentialism*. Montreal-Ithaca: McGill-Queen's University Press.

Ereshefsky, M. 2010. "What's Wrong with the New Biological Essentialism." *Philosophy of Science* 77 (5): 674–85. https://doi.org/10.1086/656545.

_____2017. "Species." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2017. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2017/entries/species/.

Fodor, J. A. 1974. "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)." *Synthese* 28 (2): 97–115.

Griffiths, P. E. 1999. "Squaring the Circle: Natural Kinds with Historical Essences." In R. A. Wilson (ed.). *Species: New Interdisciplinary Essays*: Cambridge: MIT Press: 209–228.

Hacking, I. 2015. "Natural Kinds, Hidden Structures, and Pragmatic Instincts." In R. Auxier (ed.). *The Philosophy of Hilary Putnam*. Chicago: Open Court: 337–357.

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

LaPorte, J. 2003. *Natural Kinds and Conceptual Change*. Cambridge: Cambridge University Press.

Lewens, T. 2012. "Species, Essence and Explanation." *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (4): 751–57. https://doi.org/10.1016/j.shpsc.2012.09.013.

Magnus, P. 2012. *Scientific Enquiry and Natural Kinds: From Planets to Mallards*. Houndmills, Basingstoke: Palgrave Macmillan.

Mayr, E. 1961. "Cause and Effect in Biology: Kinds of Causes, Predictability, and Teleology Are Viewed by a Practicing Biologist." *Science* 134 (3489): 1501–6. https://doi.org/10.1126/science.134.3489.1501.

Okasha, S. 2002. "Darwinian Metaphysics: Species and The Question of Essentialism." *Synthese* 131 (2): 191–213. https://doi.org/10.1023/A:1015731831011.

Putnam, H. 1975a. *Mind, Language, and Reality*. Cambridge: Cambridge University Press.

_____1975b. "The Meaning of 'Meaning.'" *Minnesota Studies in the Philosophy of Science* 7: 131–193.

Wiggins, D. 1980. *Sameness and Substance*. Oxford: Wiley Blackwell.

Wilkerson, T. E. 1988. "Natural Kinds." *Philosophy* 63 (243): 29–42. https://doi.org/10.1017/S0031819100043114.

_____1993. "Species, Essences and the Names of Natural Kinds." *The Philosophical Quarterly (1950-)* 43 (170): 1–19. https://doi.org/10.2307/2219938.

Wilson, R. A., M. J. Barker, and I. Brigandt. 2007. "When Traditional Essentialism Fails: Biological Natural Kinds." *Philosophical Topics* 35 (1/2): 189–215.

Croatian Journal of Philosophy
Vol. XVIII, No. 53, 2018

Devitt's 'Intrinsic Biological Essentialism'

URŠKA MARTINC
University of Maribor, Maribor, Slovenia

This article is about the problem of essentialism of natural and biological kinds, especially species. We will primarily focus on Michael Devitt's work "Resurrecting Biological Essentialism" (2008). We will try to prove what a good candidate for the essence of the species could be. This article puts the problem of essentialism into the context of biology and, through the usage of examples, attempts to answer that problem. We are going to try to define essentialism and determine what meaning essentialism holds in biology. We will cross-check the definitions of essentialism and compare the essence of various sciences with the suggestions of essences of species. We are going to analyse what Hilary Putnam states about natural kinds, about the so-called 'hidden structures', and what the essence of species could be. Using examples from biology, we are going to create a difference between 'underlying' and 'exterior' characteristics of organisms. We are going to analyse Devitt's 'Intrinsic Biological Essentialism' (2008) and check its advantages and disadvantages. Using examples from biology and using the analogy of examples from chemistry and biology, we will show whether Devitt's 'intrinsic biological essentialism' is valid or not.

Keywords: Michael Devitt, essentialism, natural kinds, species.

1. *Introduction*

One of the more important discussions among biologists and philosophers of biology is about the nature of species. Okasha says that the debate about the nature of species is still discussed and there is no agreement about this issue (Okasha 2002: 191).
Bird and Tobin state about essentialism:

> Kripke (1971, 1972) and Putnam (1975a) use animal kinds as examples of natural kinds for which a posteriori essences can be found. There is some implication that these essences are microstructural, intrinsic properties,

> which will be, of necessity, individually necessary and jointly sufficient for an entity to be a member of a kind. However, if species are individuals, then it is not true that species may be individuated on the basis of the intrinsic properties of their members. /…/ According to the BSC, for example, membership of a species depends on relational properties, such as membership of a certain population and interbreeding. Alternatively, the PSC refers to shared descent (Bird and Tobin 2017).

What makes this specific kind unique and not any other is the main question. Putnam says that the essence of lemon lies in its 'genetic code', and this is what makes this lemon—the individual specimen of lemon, a member of a species—lemon (Putnam 1975: 239–240).

When it comes to the 'genetic code', Putnam states that:

> At the same time the sense in which to be a lemon something has to have the genetic code of a lemon is not the same as the technical sense (if there is one, which I doubt). The technical sense, I take it, would be one in which 'lemon' was synonymous with a description which specified the genetic code. But when we said (to change the example) that to be water something has to be H_2O we did not mean, as we made clear, that the speaker has to know this. It is only by confusing metaphysical necessity with epistemological necessity that one can conclude that, if the (metaphysically necessary) truth-condition for being water is being H_2O, then 'water' must be synonymous with H_2O—in which case it is certainly a term of science. And similarly, even though the predominant sense of 'lemon' is one in which to be a lemon something has to have the genetic code of a lemon (I believe), it does not follow that 'lemon' is synonymous with a description which specifies the genetic code explicitly or otherwise (Putnam 1975: 240).

One of the problems "is that a natural kind may have abnormal members" (Putnam 1975: 140). As an example, Putnam gives the green lemon, which is still a member of the kind lemon

He writes:

> The supposed 'defining characteristics' of lemons are: yellow colour, tart taste, a certain kind of peel, etc. Why is the term 'lemon' not definable by simply conjoining these 'defining characteristics'? The most obvious difficulty is that a natural kind may have abnormal members. A green lemon is still a lemon—even if, owing to some abnormality, it never turns yellow. A three-legged tiger is still a tiger. Gold in the gaseous state is still gold. It is only normal lemons that are yellow, tart, etc.; only normal tigers that are four-legged; only gold under normal conditions that is hard, white or yellow, etc. (Putnam 1975: 140).

Putnam says that the existence of certain characteristics represents the main essential element that is common to other members of a particular natural kind (Putnam 1975: 140–141). He states:

> If I describe something as a lemon, or as an acid, I indicate that it is likely to have certain characteristics (yellow peel, or sour taste in dilute water solution, as the case may be); but I also indicate that the presence of those characteristics, if they are present, is likely to be accounted for by some 'essential nature' which the thing shares with other members of the natural kind. What the essential nature is is not a matter of language analysis but of scientific theory construction; today we would say it was chromosome

> structure, in the case of lemons, and being a proton-donor, in the case of acids (Putnam 1975: 140–141).

Can we, therefore, conclude that intrinsic characteristics are the main essential elements of a natural kind? As Putnam claims, in some cases that is true. Can the same be claimed for the species, for those organisms which are known for always evolving? We are going to try to analyse this viewpoint.

We will help ourselves with Putnam's example of the atomic number of gold (Putnam 1975).

However, let us first look at what Okasha writes: "Thus science has taught us that the real essence of gold is 'having atomic number 79', according to Kripke and Putnam—this is the underlying microstructural property which explains why all samples of gold are shiny, yellow and malleable" (Okasha 2002: 194–195).

As Putnam conclude: "To sum up: if there is a hidden structure, then generally it determines what it is to be a member of the natural kind, not only in the actual world, but in all possible worlds" (Putnam 1975: 241–242).

There is also a problem about how we classify objects. As Okasha states:

> One possible source of philosophical opposition to a relational taxonomy stems from a general view about the purpose of classifying in science. /…/. That is a philosophical commonplace. /…/ According to a widely held view, what makes one classificatory scheme more fundamental than another is that it permits more predictively useful generalisations to be formulated (Okasha 2002: 207).

Okasha states that Kripke and Putnam were not incorrect when saying that morphological characteristics are important when defining species and these criteria are "indicative of something deeper" (Okasha 2002: 203). Furthermore, Okasha says that: "… their error lies only in a mistaken view of what that 'something deeper' is" (Okasha 2002: 203).

We are going to provide an answer regarding essentialism when considering species; that is why we will start with an analysis of this problem, and then immerse ourselves into Devitt's 'intrinsic biological essentialism'.

2. *Essentialism*

> "Essentialism about species is today a dead issue"
> (Sober 1994: 163).

One of the problems of philosophy of biology is the problem of essentialism. Michael Devitt begins his own article with the same quotation from Sober. However, is this point of view valid?

Devitt explains that the term 'essentialism' is unsuitable for biologists because of its connection with "Aristotelian metaphysics" (Devitt 2008: 347).

Devitt in his work says:

> But the essentialism I have defined need not come with those Aristotelian trappings. Many philosophers would be similarly reluctant because the term 'essentialism' strikes them as quaintly old-fashioned, scholastic, even unscientific. But such reluctance would be a merely verbal matter (Devitt 2008: 347).

When we talk about 'essentialism' it is important to know that the problem of 'essentialism' will remain independent of the term.

Devitt states:

> It is the issue of in virtue of what an organism is a member of a certain Linnaean taxon; the issue of what makes an organism a member of that taxon; the issue of the very nature of the taxon. I stick with 'essentialism' because it is the term that philosophers of biology use for the doctrine that they want to reject and I want to promote. Those who are offended by the term, should replace it with one of the other ways of characterizing the issue (Devitt 2008: 347–348).

Nevertheless, we have still not defined what essentialism is.

Let's look what Robertson and Atkins says:

> Essentialism in general may be characterized as the doctrine that (at least some) objects have (at least some) essential properties. This characterization is not universally accepted, but no characterization is; and at least this one has the virtue of being simple and straightforward. (Robertson and Atkins 2018)

Furthermore, Sober claims that essentialism: "… is a standard philosophical view about natural kinds. It holds that each natural kind can be defined in terms of properties that are possessed by all and only the members of that kind" (Sober 2000: 148).

As already mentioned, one of the examples stemming from biology is the case of lemon, which was provided by Putnam. Here we can observe that he is suggesting that the essence of a species, in his case lemons, is some inner structure of an organism (Putnam 1975: 140–141).

We find it interesting what Ney says about essences:

> A historically interesting position in metaphysics is that objects have certain properties that hold of them necessarily, so-called essential properties or essences. /…/ [However,] more controversial is the issue of whether material objects like tables, chairs, or organisms have essential features (Ney 2014: 193).

Turning back to Putnam, he argues that gold is of great significance for us and we can use gold in different ways and for different purposes (Putnam 1979: 227). For example, my earrings are made of gold, the Golden Buddha statue is made of gold and there is also a champagne with flakes of gold floating in it.

Sober says that the members of the same kind have to share common properties that are characteristic of that kind, and thus what is important is not the common history of individuals but their similarity (Sober 2000: 148).

Putnam and Sober, show in their examples that the essence of gold is its atomic number. The important question for us is, therefore: What are the essential properties of species?

We could start with the example put forward by Sober. "We now can examine the idea that evolutionary theory refutes essentialism as a view about species" and that:

> One argument goes like this:
> Natural kinds are immutable.
> Species evolve.
> Hence, species are not natural kinds. (Sober 2000: 149).

The objects made of gold, for example, gold earrings, can change while "the nature of gold" remains the same way, this holds, *mutatis mutandis*, for species (Sober 2000: 149). But as Sober further states:

> Once the first premise is clarified in this way, we can see that the argument is flawed. Transmutation of the elements is possible… However, this does not undermine the idea that the chemical elements have immutable essences. Likewise, the fact that a population belonging to one species can give rise to a population belonging to another does not refute essentialism about species (Sober 2000: 149–150).

We can illustrate this using speciation. Let's look what Mayr says about speciation in his work *Systematics and the origin of species, from the viewpoint of a zoologist* (1942):

> One part of the process of speciation is the establishment of discontinuities, that is, the establishment of isolating mechanisms and their perfection to the point where reproductive isolation is accomplished and the "parent species" breaks up into two or more daughter species. (Mayr 1942: 23)

Furthermore, Ereshefsky says that: "In all but a few cases, speciation is a long and gradual process such that there is no principled way to draw a precise boundary between one species and the next" (Ereshefsky 2017).

Sober explains why essentialism is the "wrong approach regarding species" (Sober 2000: 151). While phenetics believe that species are characterize by "phenotypic or genetic similarities", biologists who are in favour of different concepts, for example, Mayr's biological concept or ecological-niche concept, don't think so (Sober 2000: 151).

One of the problems is that organisms of the same species could live in various habitats. As Sober argues, this means that they have to adjust to different environmental conditions. There are some dissimilarities between taxonomic ranks (species, genus, family, order) and this need further explanations (Sober 2000: 160).

Okasha agrees, saying that philosophers of biology acknowledge that "species are not individuated by essential characters" (Okasha 2002: 196).

Moving on, we are going to endeavour to prove whether a certain genetic characteristic is a candidate for the essence of a species. The question here is also whether DNA can be the essence of a species.

The member of the same species, for example, the members of species mute swan (*Cygnus olor*) have some "genetic similarities" among themselves (Okasha 2002: 197).

However, if we look at human beings:

> The vast majority of humans have 23 chromosome pairs, for example, while the primates most closely related to us normally have 24. But not all humans have 23 chromosome pairs—sufferers from Down's syndrome and other genetic diseases have additional chromosomes, but are still clearly human. As it is at the level of morphology, so it is at the chromosomal and genetic levels—species taxa are distinguished by clusters of covarying traits, not by shared essences. The idea that species can somehow be "defined in terms of their DNA" has no basis in biological fact, despite what many non-biologists appear to think (Okasha 2002: 197).

There are examples of organisms that have the same number of chromosomes and are of the same species, for instance, broccoli and cabbage. However, there are some examples where members of the same species have different numbers of chromosomes. Jack jumper ant *(Myrmecia pilosula)* have 1 or 2 chromosomes, female have 2 and males have 1, but they still belong to the same species, *Myrmecia pilosula*.

Furthermore, an example from botany shows that different species can have the same number of chromosomes. Barley (*Hordeum vulgare*) and Garden Pea (*Pisum sativum*) are two different species and they also differ in genus, but they have the same number of chromosomes.

One of the conclusions that Okasha built from his article, with which we do not agree, is that: "The anti-essentialist arguments of philosophers of biology show only that species cannot be defined by essential intrinsic properties" (Okasha 2002: 210).

We will, following Devitt, hope to show that this is not the fact. This will be done by analysing Devitt's thesis about intrinsic property, his 'intrinsic biological essentialism', and comparing it with arguments from Okasha and Sober.

3. *Devitt's 'intrinsic biological essentialism'*

> Essentialism is a thesis about what it is for an organism to be, say, a dog not a cat, not about what it is for, say, dogs to be a species not a genus.
>
> (Devitt 2008: 346)

Michael Devitt, in his article "Resurrecting Biological Essentialism" (2008), defends the claim that species and other taxa (genus, class, phylum, and so forth) "have essences that are, at least partly, underlying intrinsic, mostly genetic, properties" (Devitt 2008: 344).

We will try to analyse Devitt's view on essentialism in biology: A property P is an essential property of being an F iff anything is an F partly in virtue of having P. A property P is the essence of being an F iff anything is an F in virtue of having P. The essence of being F is the sum of its essential properties (Devitt 2008: 345).

Devitt argues that essence can have some variations. He says that:

> Essences can be fully intrinsic; for example, the essence of being gold is having atomic number 79. Essences can be partly intrinsic and partly extrinsic and relational; for example, the essence of being a pencil is partly being an instrument for writing, which an object has in virtue of its relation to human intentions, and partly having the sort of physical constitution that distinguishes it from a pen, which an object has intrinsically. Finally, essences can be fully relational and extrinsic; being Australian is probably an example because it seems that anything—Rupert Murdock, Phar Lap (a horse), the Sydney Opera House, a bottle of Penfolds' Grange, the expression "no worries mate," and so on—can have the property provided it stands in the right relation to Australia. (Devitt 2008: 345–346)

This classification, especially the 'fully intrinsic essence', can be used in biology. With Linnaean taxa, Devitt points to kinds which belong to Linnaean hierarchy. Those are kingdom, phylum, class, order, family, species, subspecies, etc. He focuses on species, but claims that essentialism concerns all taxa (Devitt 2008: 346).

For our discussion, it is important to know which definition of species is the best so far, so that we can see what characteristics are important when defining species. Mayr's definition (biological species concept) is the most widely accepted definition of species. It claims that: "Species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups" (Mayr 1942: 120).

On the other hand, Devitt presents two arguments supporting 'intrinsic biological essentialism'. One is that: "Such essential properties seem to be part of what 'genome projects' are discovering" (Devitt 2008: 351) and the second is about generalizations (Devitt 2008: 351).

Concerning generalizations, he writes:

> We group organisms together under what seem, at least, to be the names of species or other taxa and make generalizations about the morphology, physiology, and behaviour of the members of these groups: about what they look like, about what they eat, about where they live, about what they prey on and are prey to, about their signals, about their mating habits, and so on (Devitt 2008: 351).

Devitt agrees with Sober when he says that the "essence of species" should tell us the basic characteristics of its members (Devitt 2008: 353). For example, the essence of proteus (*Proteus anguinus*) should tell us the common characteristics for all its members and explain us why proteus is the way it is. *Proteus anguinus* have certain characteristics that are typical for all its members.

Let's look at what he says: The intrinsic difference explains the physiological difference. If we put together each intrinsic underlying property that similarly explains a similar generalization about a species, then we have the intrinsic part of its essence (Devitt 2008: 352).

We know that biologist classify most organisms according to their similarities. Devitt argues that "they do so partly on the assumption

that those similarities are to be explained by some intrinsic underlying nature of the group" (Devitt 2008: 352–353).

Devitt gives us another example, the example of the tiger. So the apparently superficial explanation points to the deep fact that there is something intrinsic, probably unknown, partly in virtue of which the animal is a tiger and which causes it to be striped. That something is an essential intrinsic property (Devitt 2008: 353).

In what follows, we will analyse whether some "underlying intrinsic, mostly genetic, properties" (Devitt 2008: 352) could be the essence of a species.

4. *A candidate for the essence of species*

> This was not any old molecule: DNA, as Crick and I appreciated, holds the very key to the nature of living things. It stores the hereditary information that is passed on from one generation to the next, and it orchestrates the incredibly complex world of the cell.
>
> (Watson, Berry, and Davies 2017: xi)

If DNA "holds the very key to the nature of living things" (Watson, Berry, and Davies 2017: xi) then it must show us some important information about organisms and what makes an organism belong to a particular species. This goes together with Devitt's idea of 'intrinsic biological essentialism'.

The thesis of this article regarding the essentialism of species is the following: we agree that one of the essential properties regarding species could be Devitt's "at least partly, underlying intrinsic, mostly genetic, properties" (Devitt 2008: 344).

The ideas supporting this thesis are as follows. 1. The first idea is connected to Devitt's argument about "genome project", which is still in progress (Devitt 2008: 351). We will be using some selected authors from biology to make our case.

Hebert et al. write:

> Genomic approaches to taxon diagnosis exploit diversity among DNA sequences to identify organisms (Kurtzman 1994; Wilson 1995). In a very real sense, these sequences can be viewed as genetic 'barcodes' that are embedded in every cell" (Hebert, Cywinska, Ball, and deWaard 2003: 313).

This procedure of recognizing species has some restriction:

> First, both phenotypic plasticity and genetic variability in the characters employed for species recognition can lead to incorrect identifications. Second, this approach overlooks morphologically cryptic taxa … Third, since morphological keys are often effective only for a particular life stage or gender, many individuals cannot be identified. Finally, although modern interactive versions represent a major advance, the use of keys often demands such a high level of expertise that misdiagnoses are common (Hebert, Cywinska, Ball, and deWaard 2003: 313).

Let us examine the following idea. 2. We want to show that the exterior characteristics of an organism, the so-called morphological marks, are not always reliable when identifying organisms.

Some unrelated organisms could be quite similar. For instance, the snake—the aesculapian snake and a lizard—the slow worm. To this very day, we still come across misconceptions that the slow worm is a snake because of the similar characteristics that these two organisms share. Most lizards have legs, whereas slow worms do not have them, thus slow worms look like snakes to laymen.

The slow worm (*Anguis fragilis*) is a legless lizard. "There are two properties by which we can distinguish the slow worm from a snake: the slow worm is able to blink and shed its tail." (*Plazilci Slovenije-jih poznamo* 2018).

With this example, we are trying to show that the exterior characteristics of organisms cannot be essential, as they could change due to various circumstances. Therefore, as this example has shown, the exterior marks might be misleading. We agree with the notion that the identification on the grounds of morphology is very important, and we do not want to devalue this particular method, but is seems obvious that only morphology is not sufficient. What follows is the examination of the taxonomic classification of the aesculapian snake and the lizard slow worm.

Kingdom: *Animalia*
Phylum: *Chordata*
Subphylum: *Vertebrata*
Class: *Reptilia*
Order: *Squamata*
Suborder: *Ophidia*
Family: *Colubridae*
Genus: *Zamenis*, Fitzinger, 1833
Species: *Z. Longissimus,* Laurenti, 1768

Kingdom: *Animalia*
Phylum: *Chordata*
Subphylum: *Vertebrata*
Class: *Reptilia*
Order: *Squamata*
Suborder: *Sauria*
Family: *Anguidae*
Genus: *Anguis*, Linnaeus, 1758
Species: *A. Fragilis*, Linnaeus, 1758
(Kryštufek and Janžekovič 1999)

As we can see from the above classification, not only do the organisms belong to different species, they also belong to different families, and, taking it even further, according to the Linnaean hierarchy, also to different suborders. If morphological marks are not essential for a species, then the essence must be something else.

Another example is variation among species. Map butterfly (*A. Levana*) have the very interesting characteristic. His wing colour and also pattern varies and these variations are dependant on the time of their birth. Those who are born in summer have a different colour than those who are born in spring (Fric and Konvička 2002: 1018).

The examples chosen are relatively simple. We can quite easily distinguish the slow worm from a snake. There are also organisms where identification according to morphological marks is almost impossible. Such species are the so-called cryptic species. Mayr defines that: "these species show the same genetic, behavioral, and ecological differences from traditional species as do phenotypically different species but do not possess the traditional taxonomic differences (Mayr 2001: 182).

That is why natural kinds must have properties that are not shared with other kinds, and this, according to the example of cryptic species (at least what exterior marks is concerned), seems not to be true.

For example, cryptic species are "very common in mycology" (Piškur 2010: 341). If these properties could be shared with other species, then these properties cannot be essential for the species.

Let's look what Piškur says:

> To unravel the morphologically identical or the hard to discern complex of species we use an approach that is not based only on the comparison of nucleotide sequences of one region (e.g. of the rDNA region), but also incorporates analyses of other sections of the genome (for example, genes for elongation factor 1-a, calmodulin, -tubulin). (Piškur 2010: 341)

Devitt suggests a certain 'underlying intrinsic' properties of organisms as the essence of a species, and at the same time states that this could be the genetic property of an organism. Therefore, we can claim that the genetic properties of organisms are a good candidate for the essence of a species, since, as it is demonstrated by the example with snake and lizard, the exterior marks of an organism may be misleading when we want to identify organisms.

Let's try to verify Devitt's thesis by using an example from the field of chemistry, i.e. an example regarding the essence of chemical elements.

The number of protons is that which determines the chemical properties of an element. So we are trying to find something that can identify a species and at the same time fulfils the same conditions as the essence of chemical elements: it provides essential properties.

Analysing an example from biology that can substantiate this thesis:

According to the American Museum of Natural history scientists compare the DNA between organisms:

> The chimpanzee and another ape, the bonobo, are humans' closest living relatives. These three species look alike in many ways, both in body and behavior. But for a clear understanding of how closely they are related, scientists compare their DNA, an essential molecule that's the instruction manual for building each species (American Museum of Natural History 2018)

When scientists analyse DNA sequences of different organisms they also learn about how organisms are related, which gives us more information about the relationship than about morphology (Murnaghan 2018).

Some research shows that different models can help us identify species. "A model COI profile, based upon the analysis of a single individual from each of 200 closely allied species of lepidopterans, was 100% successful in correctly identifying subsequent specimens" (Hebert, Cywinska, Ball, and deWaard 2003: 313).

5. *Conclusion*

In this article, we tried to analyse Devitt's 'intrinsic biological essentialism'. The problem of essentialism regarding species remains one of the central topics in philosophy of biology. If we can recognize that kinds in sciences, such as chemistry, have an essence or essential properties, then this recognition could be translated into other scientific fields—in this case, biology. We leaned on Devitt's idea that species "have essences that are, at least partly, underlying intrinsic, mostly genetic, properties" (Devitt 2008: 344). If we can determine essence in chemical elements, then we can determine essence in species. The fact is that due to the increasing advancement in technology more focus is put on genetics and genetic research, and this determines whether organisms are related or not. When genetic research shows that organisms are related, this means that they have some characteristics that all members of the group share. These characteristics are probably one of the essential properties of organisms, because these properties define a species and classify the organisms to that particular species. If essential properties are still unknown to us that does not mean that they do not exist. There remains a possibility that due to advancements in technology we will be able to find them and this could either confirm or reject Devitt's idea about essence. Nevertheless, the already mentioned idea that species have essence which is based on genetic properties seems to have the most merit.

Acknowledgments

For comments and advice, I thank Prof. Bojan Borstner and Prof. Nenad Miščević. Still it goes without saying that any mistakes in the text are exclusively mine.

References

American Museum of Natural History. *American Museum of Natural History*. Accessed February 26th, 2018 via https://www.amnh.org/exhibitions/permanent-exhibitions/human-origins-and-cultural-halls/anne-and-bernard-spitzer-hall-of-human-origins/understanding-our-past/dna-comparing-humans-and-chimps/.

Bird, A. and Tobin, E. 2017. "Natural Kinds." In E. N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition). URL = <https://plato.stanford.edu/archives/spr2017/entries/natural-kinds/>.

Devitt, M. 2008. "Resurrecting Biological Essentialism." *Philosophy of Science* 75: 344–382.

Ereshefsky, M. 2017. "Species." In E. N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition). URL = <https://plato.stanford.edu/archives/fall2017/entries/species/>.

Fric, Z. and Konvička, M. 2002. "Generations of the polyphenic butterfly Araschnia levana differ in body design." *Evolutionary Ecology Research* 4: 1017–1032.

Hebert, P., D., N., Cywinska A., Ball S. L., and deWaard J. R. 2003. "Biological identifications through DNA barcodes." *Proceedings of the Royal Society B* 270: 313–321.

Kryštufek, B., and Janžekovič, F. 1999. *Ključ za določevanje vretenčarjev Slovenije*. Ljubljana: DZS.

Mayr, E. 2001. *What Evolution Is.* New York: Basic Books.

Mayr, E. 1942. *Systematics and the origin of species, from the viewpoint of a zoologist.* Cambridge: Harvard University Press.

Murnaghan, I. 2018. *Evolution of DNA.* Accessed February 26th, 2018 via http://www.exploredna.co.uk/evolution-dna.html.

Ney, A. 2014. *Metaphysics. An Introduction.* New York: Routledge.

Okasha, S. 2002. "Darwinian Metaphysics: Species and the Question of Essentialism." *Synthese* 131: 191–213.

Piškur, B. 2010. *Kaj se dogaja z jesenom pri nas? – Četrto nadaljevanje. Gozdarski vestnik* 68 (5/6): 340–344.

Plazilci Slovenije – Jih Poznamo. *Slepec.* Accessed March 1st 2018 via https://sites.google.com/site/plazilcislovenijejihpoznamo/nasi-plazilci/slepec.

Putnam, H. 1975. "Mind, Language and Reality." In his *Philosophical Papers, Volume 2.* Cambridge University Press.

Robertson, T. and Atkins, P. 2018. "Essential vs. Accidental Properties." In E. N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition). URL = <https://plato.stanford.edu/archives/spr2018/entries/essential-accidental/>.

Sober, E. 2000. *Philosophy of Biology. Second edition.* Boulder: Westview Press.

Sober, E. 1994. "Evolution, Population Thinking, and Essentialism." In E. Sober (ed.). *Conceptual Issues in Evolutionary Biology, 2nd* edition. Cambridge: MIT Press: 161–189.

Watson, J. D., Berry, A., and Davies, K. 2017. *DNA: The Story of the Genetic Revolution. Second edition.* New York: Alfred A. Knopf.

Croatian Journal of Philosophy
Vol. XVIII, No. 53, 2018

Structured Propositions, Unity, and the Sense-Nonsense Distinction

OCTAVIAN ION

Back in the Good Old Days of Logical Positivism, theories of meaning were part of a normative project that sought not merely to describe the features of language and its use, but so to speak to separate the wheat from the chaff. In this paper, I side with Herman Cappelen (2013) in thinking that we need to rethink and reintroduce the important distinction between sense and nonsense that was ditched along with other normative aspirations during Logical Positivism's spectacular demise. Despite this, my delineation of the bounds of sense is different from Cappelen's. One of my goals in the present paper is to argue that category mistakes are paradigmatic examples of nonsensical sentences. To this end I describe one candidate for what it might be that makes category mistakes nonsensical.

Keywords: Content, category mistake, nonsense, proposition, unity, state of affairs.

0. *Structured propositions, unity, and the sense-nonsense distinction*

Language allows for the construction of meaningful units of information, but it also allows for the construction of units that according to some do not warrant the label of meaningfulness. Here might be an instance: "Streweebles floombada sharmavikssy." This sentence (if one is liberal enough to call it that) is nonsensical, since I just strung together letters into compounds without assigning them any meaning. This is not to say that the string I produced cannot be used for certain purposes. I may use it as my banking password, or use it to encode a message, etc. Despite these non-linguistic uses, it is standard practice to call such gibberish nonsense *qua* linguistic. Yet, the word 'nonsense' has been used to cover far more ground than that. Logical positivists, for instance, held that vast swathes of discourse, including almost all philosophy, contain mostly nonsense.

My concern in what follows is with how we might conceive of nonsense and its limits, and in particular with investigating what kinds of nonsense there are and what kind of failures they involve. My focus will be on grammatical nonsense, i.e. sentences that conform to the rules of grammar, to the exclusion of sentences containing mere gibberish, either in full or in part. With respect to the former, two candidates have traditionally been distinguished: sentences containing one or more terms lacking a definite semantic content and sentences containing so-called category mistakes.[1] I will follow some recent usage in calling these two candidates respectively Type I and Type II.[2] The question of whether there is Type I nonsense is, in a sense more direct: If a sentence contains one or more terms that are semantically defective, then the resulting sentence itself is rendered defective, on the popular assumption that sentential meanings are determined by the meanings of their parts and how they are combined. The question that will preoccupy us more centrally here is whether there is Type II nonsense, i.e. whether grammatical sentences ever count as nonsense. Here the way to proceed is less obvious since the sentences contain only meaningful parts arranged in ordinary ways.

To answer the question about Type II nonsense in the positive is to go against the grain.[3] I argue that category mistakes are nonsensical (section 1) and provide an account of what this means within a theory of content (section 2). In sections 3 and 4 I discuss some of the challenges this view faces, and argue that they fail to undermine it, though some serious worries loom on the horizon.

1. *Of Reactionaries and Revolutionaries*

One important nicety that needs to be settled at the outset concerns the term "meaningful", and what this term primarily applies to (in its semantic use). According to some interpreters of Wittgenstein's *Tractatus* (and, in this intensional context of their theorizing, according to early Wittgenstein himself) all forms of linguistic nonsense, i.e. gibberish, category mistakes, sentences containing terms that lack meaning, all suffer from a common failing: In each instance, the speaker has failed to assign meanings to their terms. Following the nomenclature of its proponents, I'll call this position Austerity.[4] Not everyone is an Austerity theorist. To say that gibberish and sentences containing terms without content and category mistakes are all nonsensical is not (automatically) to say that there is a unique sort of failure at work in each.

[1] See, for instance Carnap (1932), although Carnap himself did not use the term 'category mistake.'

[2] Some theorists call these different types Type I and Type II nonsense respectively. See Conant (2002: 380–383); and Cappelen (2013).

[3] Theorists who deny that Type II nonsense exists include Lambert (1969); Camp (2004); Magidor (2009; 2013).

[4] See Conant (2002: 380).

Nonsensicality might after all turn out to admit of different species. Austerity requires the following further commitment which I'll abbreviate SF for Sentence First:

(SF) The parts of a sentence are meaningful only insofar as the sentence of which they are parts is meaningful.[5]

The motivation for (SF) is that, prior to grasping the meaning of the entire sentence, we cannot determine what contribution the individual sub-sentential expressions make. As a result, unless the sentence as a whole has a definite meaning, no part of it does, since the parts are meaningful only insofar as they contribute to the meaning of the whole. Most meaning theorists however do not accept the principle that the meaning of the whole determines the meaning of the parts. Rather, most theorists take sub-sentential meanings to determine the meanings of sentences, thereby accepting some principle like the Compositionality Principle:

(CP) The meaning of a sentence depends on the meanings of its parts and how they are composed.

Given the ubiquity of CP (in its various formulations), I will not attempt a defense of it, but rather take it for granted here. What follows, once we accept CP?

Take some simple sentence of the form [a is F], composed out of a proper name "a" and a predicate, "F". What divides philosophers writing on Type II nonsense is the question whether all such predications are meaningful when grammatical or whether there is some further constraint on the kinds of terms that can replace "a" and "F". One camp claims that all such sentences are meaningful. The other camp claims that replacements for "a" and "F" must respect category boundaries.[6] I'll call the former *Revolutionaries* and the latter *Reactionaries*.

Reactionaries believe that linguistic strings that cross categories are nonsensical. According to them, saying of a flesh and blood individual, like Caesar, that he is a prime number or false or bijective is nonsense, meaningless, or not significant. My aim in this paper is to advance the Reactionary thesis that cross-categorial strings are indeed nonsensical, and specifically to do so by explicating what semantic defect is exhibited by such sentences. Once I state this view in the next section, I discuss one line of opposition recently advanced by Elisabeth

[5] I formulate SF using the term 'sentence' here, even though Wittgenstein uses 'proposition'. The reader can switch them up here if they strongly object to my formulation.

[6] Precisely what categories there are is a subject of debate. Here I will operate on the minimal assumption that there is at least one division amongst categories that must be respected, the division between the concrete and the abstract. Paradigm instances of the former include individuals, like Caesar, and locations, like Cleveland. Paradigm instances of the latter include mathematical objects, like the number 3, and subjects of study, like geography. The category errors that I am concerned with match objects in one category with predicates that apply to objects from the other category.

Camp and show that it is wanting. In the penultimate section I consider a more pressing objection raised by Ofra Magidor, although it is not clear, as I hope to show, that it is effective against the view I present.

I should clarify at the outset that my proposal is pitched at the level of content, not at the level of meaning more generally.

2. *Category mistakes as nonsense*

At the level of semantic content, one popular sort of account takes sentences to express structured propositions and takes names and predicates to have as semantic contents individuals, properties and relations. The sentence

(1) Sally is taller than John,

on such a view, expresses a proposition which can be roughly represented as

(2) <<Sally, John>, being taller than>.

Some proposition theorists extend the role they play further than this. A bit more controversially, aside from being what declarative sentences express, propositions are also taken by many to play the role of contents of propositional attitudes like believing and desiring, and referents of that-clauses.[7] The proposition expressed by a sentence is the informational content of the sentence. My Reactionary thesis (like Cappelen's and Carnap's before him) is that a sentence counts as nonsense insofar as it fails to express a proposition; a thought is nonsensical insofar as it does not have a propositional content; a 'that'-clause is likewise nonsensical if it doesn't refer to a proposition.[8] Of course, the difficulty for the line I'm taking lies in specifying how these content failures occur, and explaining the roles of cross-categorial sentences in our linguistic practice.

To say that a grammatical sentence fails to express a proposition is to say that some ingredient of its propositional content is absent. As mentioned, in the case of sentences containing terms that lack semantic content, it is easy to see why no proposition is expressed, since there is a gap in the would-be structured proposition. What about cross-categorial sentences? While it is common for content theorists to represent propositions as set-theoretic objects, it is clear that an account of the nature of structured propositions is needed to account for their

[7] For grounds for hope that propositions can play such roles see chapters 1 & 2 in King, Soames, and Speaks' *New Thinking about Propositions* (2014), for an account of the value of taking propositions to play these various roles. For grounds for despair about propositions being able to play such roles see Jubien (2001) and Weber (2012).

[8] One anonymous referee asks about contextualist views which take sentences not to express propositions (but perhaps only propositional skeletons), yet for all that are not semantically defective. However, the category mistakes that are my focus fail in expressing unified propositions across the board, regardless of differences of context or point of evaluation.

unity in a way that explains how they can play the sorts of roles they are supposed to.[9] In particular, theorists have been interested in figuring out how propositional components must be unified so as to have truth-values and to be able to represent how things stand in the world. This opens up the possibility that cross-categorial sentences might fail to express propositions not because some subsentential constituent is semantically defective, but rather because their constituents lack the requisite unity. It is this latter possibility that I explore here.

There are some caveats that need to be mentioned before proceeding. First, I do not take the term 'nonsense' to apply to sentences *qua* types, but rather to sentences-in-contexts.[10] Second, I take it to be a live possibility that a sentence can appear to an agent to have propositional content even if it doesn't, in the sense that the internalized cross-categorial sentence can function in the way that contentful internalized sentences do within an agent's cognitive economy, despite being contentless.[11] Third, in some instances, what creates this illusion of content is a tendency on the part of the interpreter towards resolving discrepancies involved in the interpretation of cross-categorial strings by supplementing and (or) otherwise modifying what is being thought so as to maximize coherence.

The picture being described here has the advantage of diagnosing a common failing as the source of both Type I and II nonsense. In this sense, it is in agreement with the Austerity thesis despite the fact that it accepts compositionality. I take it to be a welcome result that we arrive at a notion of nonsense that unites the various forms the phenomenon can take.

So far, I've suggested that the semantic contents of cross-categorial strings are not propositions because they lack unity, but theorists are divided on the issue of just how propositional constituents *are* unified. There are two main ways to conceive of propositions, either as mind-independent or as mind-dependent. The choice here hinges on whether

[9] For a detailed exposition and criticism of the content theorists, see King (2009) and Soames (2010).

[10] This is done in part to avoid issues of ambiguity, but more importantly to put aside the possibility that a sentence which appears to make sense *prior* to fixing the referents of its pronouns and demonstratives can thereby be thought as meaningful even once the references are fixed in ways that cross category boundaries. Such a strategy would be question-begging if used against an account like the one sketched in this section, where nonsense amounts (roughly) to lack of informational content.

[11] In some such cases, the thinker either embellishes the mental sentence or the uttered sentence by interpreting it in a loose way so that, ultimately, what is being assented to or asserted by the sentence involves information that the concepts deployed do not themselves semantically encode. Upon hearing, for instance, "John's toothbrush was pregnant", the hearer can embellish what is said by conceiving of John's toothbrush as come alive (as in a cartoon) and some part of this new entity being distended in the way a pregnant female's belly is. Our minds can do this, and it's great that they can, but I do not consider the embellished scenario a faithful literal interpretation of the sentence "John's toothbrush was pregnant."

one takes propositions to be representational or not. Theorists who take propositions to be non-representational identify them with states of affairs (Richard (2013); Speaks (2014)).[12] Theorists who take them to be representational identify them either with types of mental acts (Soames 2010, 2014) or with structured relations that have the extrinsic dispositional property of being cognized in certain ways (King 2007, 2009, 2014). In the rest of this section I describe a version of each of the two conceptions and discuss how they might accommodate nonsense.[13]

2.1. *Propositions as States of Affairs & Nonsense*

One way to think of the contents of sentences and propositional attitudes is as states of affairs. This view has surfaced intermittently over the last few decades,[14] however the identification of propositions with states of affairs has only received its most recent statement in Richard (2013). States of affairs are variously glossed by Richard as possibilities, ways things could be, and more specifically, as properties of situations. Richard summarizes the view as follows: "States of affairs are certain properties, ones picked out by terms of the form *the property of being a situation in which the objects o1, … , on instantiate the properties p1,…, pj in way I*" (Richard 2013: 704). On this view, the objects, properties and relations that are the semantic contents of a sentence are unified into a proposition by being parts of a state of affairs, i.e. by being parts of a complex property of situations. As Richard emphasizes, some sentences represent states of affairs that obtain ("it's raining in Chicago"), some represent states of affairs that do not obtain ("Paris is north of London"), while some represent states of affairs that could not possibly obtain ("Water is not H2O").

My contention is that on this account, category mistakes can be

[12] I do not mean to suggest that Richard and/or Speaks are the first philosophers to advocate that propositions are states of affairs. The history of the discussion of the nature of propositions has a long history. I am working with their accounts because they present the most recent formulations of this sort of theory.

[13] This division is not meant to be exhaustive, as some content theorists reject propositions altogether. See Jubien (2001), Simchen (2013), among others. It is not clear however that rejecting propositionalism requires one to adopt a revolutionary attitude towards category mistakes. As I argue in section (2.2) below, even on a no-propositions view on which every sentence has a truth-value computed on the basis of the semantic values of its component expressions, it can still be appropriate to speak of how things are represented by that sentence. This in turn requires there to be a state of affairs made up from those semantic values that either obtains or fails to obtain (either actually or necessarily). A sentence fails to express a proposition, as before, if there is no such corresponding state of affairs.

[14] Here's Fodor in an article from 1984: "The paradigmatic representation relation […] holds between things of the sorts that have truth values and things of the sorts by which truth values are determined. I shall usually refer to the latter as "states of affairs", and I'll use '-ing nominals' as canonical forms for expressing them (e.g., 'John's going to the store'; 'Mary's kissing Bill'; 'Sam's being twelve years old next Tuesday')" (233).

seen as marking a distinctive sort of failure, whereby a sentence or thought fails to represent *any* state of affairs. What this amounts to saying is that there can be no state of affairs, e.g. of Caesar's being prime. This is not a matter of there being a state of affairs corresponding to the sentence, albeit one that merely does not (or even cannot possibly) obtain. Were this so, the sentence would be false or necessarily false rather than nonsensical. Rather, category mistakes have *no correlate whatsoever* in what we might call metaphysical (or perhaps even logical) space, even when this space is expanded to include impossibilities. Why might this be?

There are two reasons for thinking that category mistakes fail to express propositions on the current view, depending on whether we focus on the relation of expressing or whether we focus on what is being expressed. The first reason is the more roundabout of the two. Recall that on the present view propositions are states-of-affairs, but to think a thought with a certain state-of-affairs as its content is to represent that state of affairs. Presumably a sentence expresses a state of affairs if it is possible (at least in principle, for someone) to represent the state of affairs that it expresses. But what is it to represent Caesar as being prime? Worse yet, what is it to represent aluminum as having divorced democracy? Here the answer that suggests itself is that the failure of a sentence S to *express* a proposition is tied to a failure on the part of language users to produce a unified thought that combines radically disparate subject-concepts and predicate-concepts.

The second and stronger reason to think there can be no states of affairs corresponding to category mistakes might be worked out by appeal to the essential properties of states of affairs.[15] It is difficult to believe that the mere fact that we can combine words and concepts in certain ways implies that there is such a thing as Caesar's being prime, or the number six's being taller than a can of soup. States of affairs are, as per Richard's definition above, are properties such as the property of being a situation in which some object instantiates some property. But it is reasonable to think that the range of situations is limited by what kinds of properties it is possible for the objects to instantiate.

Something is human (or something is a number) in virtue of the essential properties it possesses. Suppose that I think a thought corresponding to the English sentence "Caesar is baking the number 2". What state of affairs would this putative thought represent? Whatever else, the state of affairs has to be one that contains Caesar and the number two, and be such that the former stands to the latter in the relation of *baking*. But there is no such relation that holds together such relata, and hence no such state of affairs.

The obvious response to this line is to claim that what goes for "Caesar is baking the number 2" goes for all false (or: necessarily false)

[15] Westerhoff suggests this as an option in his monograph, *Ontological Categories* (2005: 94).

sentences.[16] Consider "John is taller than Sally", uttered in a context where Sally is actually taller than John. Isn't it essential to the being-taller-than relation that it holds of x and y only when x exceeds y in height? Yes, that's why the sentence is false, since Sally and John are not related in the way that the sentence says they are. Recall that, on the current picture, to say a sentence is false is to say that the state of affairs it expresses does not obtain, and so we are saying *of a certain relation* that it does not actually hold. Likewise, to say that a sentence is necessarily false is to say that the state of affairs it expresses cannot possibly obtain, which is not to deny that there is such a state of affairs, only that it does not obtain in any possible circumstance. The situation is different with respect to the sentence "Caesar bakes the number two" since no relation between a person and a number *could be a baking relation.* As such, the case of category mistakes can be differentiated from that of necessarily false sentences since in the latter there is a relation that is expressed albeit one that fails to hold of necessity. In the former case there is no relation that may be said to fail to hold. Relations, like objects, have essential properties (or essences). Much of the problem posed by the distinction between category mistakes and necessarily false sentences is due to the tightly bound relation between what is essential and what is necessary.

The possibility of constructing the syntactic string "Caesar bakes the number 2" creates the illusion that we are representationally carving out a possibility, when we are not actually latching onto anything at all. Contrast this with a non-cross-categorial sentence representing an impossible state of affairs, e.g. "my jacket is both completely blue and completely green". In this latter instance, a state of affairs is represented, although no situation can possibly instantiate it. The option we are currently surveying maintains that facts about the essential properties of objects and relations set the boundary for what states of affairs there can be (regardless of whether or not they obtain). Limitations on the constructability of states of affairs are plausibly seen as generated by constraints on what kinds of objects can instantiate what kinds of properties and relations. Importantly, properties specific to spatio-temporal concrete individuals cannot be ascribed to abstract objects and vice versa.

One pressing concern stems from the fact that category mistakes have their own peculiar phenomenology. There is something that it is like to process a cross-categorial sentence, a certain oddness or funniness to it. Yet this feel is often missing when we process negated or otherwise modalized cross-categorial sentences, such as "the number two couldn't have moved next door to us" or, less extravagantly, "Caesar isn't a prime number". Now, on the state of affairs theory of

[16] The first response of this sort is found in Quine's *Word and Object* (1960: 229). There, Quine argues that it is simpler to take cross-categorial strings to be false rather than nonsensical on the assumption that we cannot non-arbitrarily distinguish category mistakes from contradictions.

propositions discussed, a sentence is false (and its negation is true) at a context if the state of affairs it expresses fails to obtain there. But if cross-categorial sentences fail to express states of affairs then it is not clear how to explain the oddness asymmetry and the fact that "Caesar isn't a prime number" isn't only lacking in oddness, but strikes us as patently true. But one reason this claim strikes us as true is likely that we interpret it as saying that "prime number" is not a property that it makes sense to ascribe to Caesar, not that it is a property that Caesar lacks, like a decent haircut or a thin nose.

I've assumed throughout this discussion that states of affairs are not, contra Richard's gloss on their nature, possibilities. I needed to do so in order to draw the distinction between necessarily false sentences and category mistakes. This move is independently motivated since Richard, and presumably all other content theorists, want to hold that meaningful sentences can express propositions that could not possibly obtain (i.e. are necessarily false).

2.2. *Propositions as Types of Acts & Nonsense*

Unlike Richard, Scott Soames argues that we should think of propositions as types of cognitive acts that agents perform. More specifically, he identifies the proposition *that snow is white* with the cognitive act type of *predicating* whiteness of snow. On this picture, to say that two agents entertain the same proposition (e.g. that snow is white) is to say that they each engage in a cognitive act of predicating whiteness of snow. Soames takes predication to be a primitive, not further analyzable notion.

Unfortunately, it is not very clear how the propositions-as-act-types proposal fits in with the view that cross-categorial strings fail to express propositions. Since propositions are identified with types of acts, to say that a sentence fails to express a proposition amounts to saying that there is no type of act of the relevant sort. To return to our running example, to say that "Caesar is prime" does not express a proposition is to say that there is no type of act of predicating *being prime* of Caesar. At first glance, this result looks problematic, since it is not clear what kind of story we might run to explain why some properties are predicable of certain objects while others are not.

To see why one nonetheless might be inclined to think that cross-categorial strings fail to express propositions on this view, it is helpful to consider how Soames's account is supposed to be an improvement over the view that propositions are set-theoretic objects. The central reason, according to Soames, is that the account is able to accommodate what many take to be a central feature of propositions, namely that they represent. What makes the type of act of predicating whiteness of snow a proposition is that in performing instances of that act type agents *represent snow as white*. This suggests a natural reason for thinking that cross-categorial strings do not express propositions since

(presumably) agents do not or *can*not represent e.g. Caesar *as* prime or *as* a bijective function or *as* baking numbers. Agents cannot represent these things because these are not ways that Caesar can be. At the very least, it must be clarified what it means to say that someone represents Caesar in these ways. Baking a cake, for instance is an activity. If someone says or thinks, "Joe is baking gingerbread cookies," that person represents Joe as engaging is a certain sort of activity. But does this commit us to claiming that thinking "Caesar is baking the number three" also involves representing Caesar as engaging in an activity, perhaps a quixotic one? That is as least far from obvious. The very notion of representing presupposes that something is being represented, e.g. the act of predicating whiteness of snow is an instance of a proposition because in this act snow is represented as being white. While this does not assume that it is either physically or even logically possible for snow to be white, (it can be meaningful to talk about impossibilities after all as mentioned in the previous subsection), what it does assume is that there is a way in which things are characterized as being. That is to say, it assumes that snow's being white is a state of affairs. From a slightly different angle, to say that an agent represents o as F requires us to assume that there is some standard of correctness that a representation can approximate to a greater or lesser degree. However in the cases we are considering here we are not dealing with ways that the individuals under consideration can be.

Ultimately, whether we take propositions to be states of affairs, or whether we take them to be types of acts, there must be such states in all cases where something is represented as being a certain way. We cannot represent a as F unless there is such a thing as *a's being F*. For reasons discussed in the previous subsection, this creates an opening for how we might understand the content failures that are specific to category mistakes.

It has been the burden of this section to show that failure to express a unified proposition might provide a reasonable way to distinguish between what counts as nonsense and what doesn't. I turn now to consider some reasons Elizabeth Camp has offered for resisting the Reactionary thesis and for adopting a Revolutionary attitude instead.

3. *Camp and the Generalized Generality Constraint*

Camp argues that cross-categorial strings (i.e. category mistakes) are not nonsensical as Reactionaries would have it. On her view, what makes such strings significant is that they "can express thoughts; and competent thinkers both are able to grasp these and ought to be able to" (Camp 2004: 209). In particular, what such cross-categorial strings express are propositional thoughts, which are composed out of concepts, with both the thoughts and their constituent concepts being individuated by their possession-conditions. Furthermore, the conditions for both concept and thought possession are cashed out in terms

of the inferences such concepts and thoughts can participate in. As she puts it, "[part] of what it is for someone to possess a concept, on this view, is for that concept to be fully caught up in a network of potential thoughts—for it to combine generally with the thinker's other concepts (subject, that is, to a mental analogue of being syntactically well formed)" (Camp, 2004: 210). By focusing on the conceptual level Camp turns the discussion of the meaningfulness of cross-categorial strings into a discussion of the acceptability of Gareth Evans' Generality Constraint:

(**GC**) If a subject can be credited with the thought that a is F, then he must have the conceptual resources for entertaining the thought that a is G, for every property of being G of which he has a conception.

Now Evans himself was not a Revolutionary in the sense presently under discussion and, as Camp notes, he qualifies (GC) in a footnote, by adding "with a proviso about the *categorical appropriateness* of the predicates to the subjects" (Evans as cited in Camp 2004: 212).[17] It is precisely this proviso that Camp aims to undermine.

Camp's argument for the unconstrained generality of (GC) relies on thinkers' capacity to deploy cross-categorial strings in material reasoning by drawing inferences to and from them. This capacity, by itself, she takes to provide evidence of the meaningfulness of such strings. Among her examples, she considers the following inferences from (3) to (4) and (3) to (5):

(3) Caesar is a prime number.
(4) Caesar lacks efficacy.
(5) Caesar could not be an effective emperor.

Neither of these last two sentences is a cross-categorial string, which thwarts the objection that all such inferences might suffer from the same "categorical inappropriateness" as the original from which they are inferred. Nor are the inferences purely formal, rather they are material. String (4) follows from (3) only given some knowledge of what it means to be a prime number, unlike

(6) Something is a prime number.

Nevertheless, Camp thinks that the material inferences only achieve their desired effect (of displaying the meaningfulness of (3)) once we take a bit *more* on board than pure inferentialism.[18] In her words "[it] does seem that some referential component is also essential for full understanding" (Camp 2004: 222). Precisely what the nature of this referential component might be however is left unclear. She continues,

> But I need not hold the view that grasping inferential role is all there is to concept possession. By hypothesis, the thinker under consideration, because supposedly otherwise competent with respect to the constituent con-

[17] For the original citation see Evans (1982: 101).

[18] Where pure inferentialism presumably is the view that the meaning of some thought [t] is constituted entirely by its inferential liaisons.

> cepts, does meet any such additional requirements for concept possession. (Camp 2004: 222)

To sum up, cross-categorial strings should be counted as significant, according to Camp, insofar as they possess substantial inferential roles, and speakers actually make use of these inferential roles, for instance when processing metaphors.

How should the Reactionary respond? One general worry about Camp's strategy is that there can be strings which agents internalize and from which they can draw seemingly reasonable inferences *even in cases where the string in question is itself nonsensical* on account of containing terms that lack a unique semantic content relative to their linguistic community. That is to say, as long as there exists Type I nonsense, and Camp concedes this much, than her argument from material inferences would reveal those sentences to be meaningful. If so, then it simply isn't true that the inferential liaisons of a putative thought suffice for rendering it meaningful.

More centrally, there appears to be a significant tension internal to Camp's argument. As mentioned, she does not subscribe to a pure inferentialism, which she rightly fears would not allow for conceptual mastery. This is not supposed to raise any serious worry however, since Camp tells us that the agents she is considering possess *standard conceptions* of the terms in cross-categorial strings. The problem is that Camp seems to want to use inferentialism for the purpose of allowing the inferences in question to give meaning to the cross-categorial strings, yet those very inferences would only be drawn by an agent if the agent's conception of the referent (e.g. "Caesar", "prime number") is non-standard or doesn't come into play at all.[19] Consider for instance just one of the resources Camp draws on: Someone can meaningfully ask, "Is Caesar a prime number?" But this cannot serve Camp's intended purpose of testifying to the meaningfulness of "Caesar is a prime number" because only someone lacking a standard conception of "Caesar" would sincerely ask the question. What this shows is that there is a tension between being conceptually competent and accepting cross-categorial sentences. Again, consider the slightly more elaborate case from above, i.e. the inference from (3) to (4):

(3) Caesar is a prime number.
(4) Caesar lacks efficacy.

Camp claims that being able to draw such inferences is required for being fully competent. To the contrary, I claim that drawing such inferences reveals that one does not possess some of the relevant concepts. Inferring from (3) to (4) has absolutely nothing to do with Caesar. If the agent does possess a standard conception of Caesar, it should be pretty clear that the inference from (3) to (4) is itself a purely formal manipulation. My contention is that there is no foreseeable way to re-

[19] Or at least so weak that they merely mimic concept possession.

solve this tension in a non-question-begging way, i.e. in a way that does not ultimately either resort to pure inferentialism or to reliance on non-competent language users.

4. *Magidor's Defense*

In this section I take up the task of critically assessing some of the central arguments of the most resolute Revolutionary in recent literature, Ofra Magidor. While she advances numerous arguments in the revolutionary vein, only two of these concerns precisely the sort of content-centered account that I have been discussing. I argue that none of those arguments provide reasonable grounds for taking category mistakes to be meaningful. My reply to the first argument however is indirect since I rely on the negative consequence that Magidor's argumentative strategy would preclude the possibility that there is Type I nonsense not just Type II.[20] Let's turn to those arguments now.

4.1. *Argument from Propositional Attitude Ascriptions*

The first argument that Magidor advances against content theories relies on the fact that category mistakes can occur in propositional attitude ascriptions, one of her examples being, "John believes that the theory of relativity is eating breakfast." This sentence appears meaningful, and if it is, then it must be composed of meaningful parts. Since "the theory of relativity is eating breakfast" is part of the sentence, it follows that *it* too is meaningful. *Prima facie*, this is a strong argument. My strategy for defusing it is to argue, as I did in response to the synonymy argument, that it is too strong.

Consider a pair of empty terms, "amphidentric" and "quarthidentric" that I make up in order to teach my students an important lesson about language. Suppose that, in a lecture, I deploy these terms many times, and dramatically argue that things that are amphidentric cannot possibly be quarthidentric. After class is over, Sam says to Jordan: "Wow! Our prof really believes quite strongly that nothing can be both amphidentric and quarthidentric." Surely we understand, in some sense what Sam *said*. Nevertheless, the 'that'-clause Sam used fails to express a proposition, and thereby is not truth-evaluable. I think this suffices to show that the argument from propositional attitude ascriptions is suspect. Not all that-clauses need to express propositions, some can merely appear to do so.

[20] A key target here is Herman Cappelen who accepts Type I nonsense while also claiming to be convinced by Magidor's arguments that category mistakes are meaningful (Cappelen 2013).

4.2. *Argument from Metaphor*

Magidor's final argument is to claim that processing the literal meaning of many category mistakes is necessary because they are metaphors. Her strategy is to divide up theories of metaphor into ones which require category mistakes to be meaningful and those which do not and show that the former candidates are more promising than the latter. The theories that she takes to be most successful are Gricean theories which assimilate metaphorical meaning to conversational implicatures and Davidson's non-cognitivism which denies metaphorical meanings exist. Grasping the literal meaning of a cross-categorial metaphor is crucial, on Gricean theories, in order to commence the process of exploring alternative implied meanings that could be intended. Literal meanings are crucial on the Davidsonian picture, because the relevant sentences do not have any other meanings. While I think that metaphors pose an interesting and important problem, I do not think that the argument in any way decisively rules out the Reactionary account.

To see why, let's consider as an example the metaphor that forms the core example of Lepore and Stone's discussion in chapter 4 of their recent book *Imagination and Convention*: "Love is a snowmobile racing through the tundra" (2015). What generates the metaphorical meaning here? If we suppose, with Lepore and Stone that uses of metaphorical language prompt analogical thinking, the question remains open as to whether this process begins with processing of the literal meaning of the metaphor or not. For the Gricean, the hearer grasps the literal content of the utterance, namely the proposition *that love is a snowmobile racing through the tundra*. The speaker then presumably realizes that this proposition is trivially false, inappropriate, etc. As a result, the hearer infers that the speaker intended one or more other propositions to be grasped. But this is not the only way of proceeding. Another option is to hold that the hearer interprets the sentence without thereby coming to believe the proposition, i.e. despite not being able to represent love *as* a snowmobile's journey, which prompts her to search for analogies between love and a snowmobile's journey that might make sense. For the Reactionary thesis proponent while speakers do grasp a proposition, they do have a metalinguistic understanding of the cross-categorial sentence: they understand that the speaker ascribes the property of being a snowmobile to love. It is unclear why this cannot suffice in generating the search for alternatives or the analogical thinking processes. For the Davidsonian on the other hand, there are no metaphorical meanings, but this does not automatically entail that the sentence must express a proposition unless we have independent reasons to hold this view. If I am correct, then we can resist this conclusion, arguing that the sentences fail to express unified propositions.

4.3. *Argument from Partial Propositions*

A more pressing worry stems from a very different sort of argument Magidor launches in the fourth chapter of her book (2013: 81–89). Here is the structure of her argument there:

i. Either we hold that category mistakes express propositions that are truth-valueless, or we hold that they fail to express propositions and are truth-valueless.
ii. If we hold the second of these positions (as the account I have embraced requires), then we are committed to the existence of partial propositions, propositions that are true at some worlds and lack a truth-value at others. She uses as example the sentence "The thing I am thinking of is green," where the definite description is understood to function as a non-rigid singular term. Now, while in actuality the definite description picks out a table, in some other possible world w* it may very well pick out the number two, since that is the thing I am thinking of there.
iii. But Magidor deploys a modalized version of an argument advanced by Williamson[21] to show that a proposition cannot lack a truth-value at any possible world.
iv. But then w* cannot be a world where the proposition is neither true nor false. So the proposition must be truth-valued even in w*

One thing to note about this argument is that it might have independently unpalatable consequences if it generalizes to the case of contingent semantic paradoxes, as Magidor notes in her footnote 15 (2013: 88). Another option that might be worth pursuing is to deny that the view I defend commits me to accepting that claim that the proposition expressed by "The thing that I am thinking of is green" is truth-valueless at w*. What Magidor shows is that the modalized version of The Williamson argument requires us to deny that a proposition can fail to have a truth-value at some possible world, but this is not the situation with w*, since w* is not a world where the proposition fails to have a truth-value, but rather a world where it fails to be a proposition.

5. *Conclusion*

The focus of the paper has been on the most worrisome candidate for qualifying as nonsense, namely category mistakes. I have attempted to show that there is room for the view that such sentences are indeed nonsensical, and described the kind of content-failure that they might

[21] The argument as stated in Magidor's *Category Mistakes* goes as follows (2013: 87):
(NT) Necessarily, the proposition that p is true if and only if p.
(NF) Necessarily, the proposition that p is false if and only if not p.
(1) Possibly, the proposition that p is not true and the proposition that p is not false.
(2) Therefore, (2) Possibly, not p and not not p.

involve, and attempted to fend off some general attacks raised against this sort of view.

References

Camp, E. 2004. "The Generality Constraint and Categorial Restrictions." *The Philosophical Quarterly* 54: 209–231.

Cappelen, H. 2012. *Philosophy Without Intuitions*. Oxford: Oxford University Press.

Cappelen, H. 2013. "Nonsense and Illusions of Thought." In J. Hawthorne (ed.). *Philosophical Perspectives 27, Philosophy of Language*: 22–50.

Carnap, R. 1932. "The Elimination of Metaphysics through Logical Analysis of Syntax." *Erkenntnis* 2: 60–81.

Conant, J. 2002. "The Method of the Tractatus." In E. H. Reck (ed.). *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*. Oxford: Oxford University Press.

Evans, G. 1982. *The Varieties of Reference*. Oxford: Clarendon Press.

Fodor, J. A. 1984. "Semantics, Wisconsin Style." *Synthese* 59: 231–250.

Jubien, M. 2001. "Propositions and the Objects of Thought." *Philosophical Studies* 104 (1): 47–62.

King, J. C. 2007. *The Nature and Structure of Content*. Oxford: Oxford University Press.

King, J. C. 2009. "Questions of Unity." *Proceedings of the Aristotelian Society* 109: 257–277.

King, J. C., S. Soames, and J. Speaks. 2014. *New Thinking about Propositions*. Oxford: Oxford University Press.

Lambert, K. 1969. *The Logical Way of Doing Things*. New Haven: Yale University Press.

Lepore, E. and M. Stone. 2015. *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford: Oxford University Press.

Magidor, O. 2009. "Category Mistakes Are Meaningful." *Linguistics and Philosophy* 32: 553–581.

Magidor, O. 2013. *Category Mistakes*. Oxford: Oxford University Press.

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

Richard, M. 2013. "What are Propositions?." *Canadian Journal of Philosophy* 43 (5–6): 702–719.

Simchen, O. 2013. "Token-Reflexivity." *The Journal of Philosophy* 110 (4): 173–193.

Soames, S. 2010. *Philosophy of Language*. Princeton: Princeton University Press.

Weber, C. 2012. "Eternalism and Propositional Multitasking: In Defense of the Operator Argument." *Synthese* 189 (1): 199–219.

Westerhoff, J. 2005. *Ontological Categories: Their Nature and Significance*. Oxford: Oxford University Press.

Croatian Journal of Philosophy
Vol. XVIII, No. 53, 2018

Do Conversational Implicatures Express Arguments?

MARTINA BLEČIĆ
University of Rijeka, Rijeka, Croatia

I suggest that the idea that conversational implicatures express argument can be significant for the notion of communicational responsibility. This underlying argument should be included in the reconstruction of conversational implicatures as a justification for the belief formed by the hearer on the basis of indirect communication. What makes this argument specific is the fact that its only explicit element is the speaker's utterance taken as its initial premise. In order to reconstruct all the other elements, the hearer has to take into consideration factors such as the context and general knowledge of the shared language and the world. As the reconstruction of conversational implicatures in general, the reconstruction of implicatures as arguments is only potential. It is proposed that we should consider conversational implicatures as reason-giving arguments in which the speaker (arguer) addresses a hearer who does not need to reply. In those cases, the speaker is not trying to convince the hearer to accept his position but is explicitly stating a reason in support of his intended message. I believe that this approach can strengthen the idea of the speaker's communicational responsibility for an implicated message even in the case when he wants to distance himself from it.

Keywords: Conversational implicature, indirect communication, arguments, argumentation, communicational responsibility, justification, rationality.

1. *Introduction*[1]

Conversational implicatures are generally seen as a cancellable pragmatic occurrence that can never convey a message with absolute certainty. Additionally, the burden of responsibility for a belief formed on the basis of an implicature is often put on the hearer. This consider-

[1] The writing of this paper was supported by the Croatian Science Foundation under grant number IP-06-2016-2408.

ation can often be found in works dealing with the possibility of indirect testimony and indirect lying (labelled "misleading" or "deception").

In the literature linking pragmatic phenomena with argumentation, the leading approach consists in adding elements from pragmatics—often speech act theory—to the theory of argumentation in order to present the process of argumentation as a specific use of language that occurs among speakers in a certain disputational context. The idea I wish to present in this paper is that treating implicatures as arguments could help us question the aforementioned characterization of implicatures and portrait them as a more robust speech phenomenon, i.e., a phenomenon that is not guided by largely unpredictable private communicational inclinations but one that can be objectively evaluated on the basis of general rational and communicative principles. I will not look into how pragmatics can help the study of argumentation. I will go the other way around and try to show how insights from the study of arguments and argumentation can help us with the understanding of conversational implicatures, especially if we wish to explore their normative dimension.[2]

2. *Conversational implicature*

I will start by presenting some simple examples of conversational implicatures, a term introduced by H. P. Grice (1989). I will not go into detail about Grice's theory and the kinds of implicatures he presents. For the readers who are already familiar with his classification it is enough to say that I will focus on particularized conversational implicatures and that my interest in this particular kind stems from the fact that they are often characterized as the one least prone to systematization. I take that they are guided by rationality and conventionality in a larger degree than it is usually believed.

The first example is the following:

> Ani: How much longer will you be?
> Ben: Mix yourself a drink.

This dialogue starts with Ani asking a question. The answer to the question should be known to the addressee, Ben, since it is directly related to him. She receives a response that can be interpreted as saying that Ben needs more time to get ready. The question is how can

[2] This is not to say that the evaluation of conversational implicatures as phenomena that express arguments is not a valuable exploration of the limits of arguments. As Goddu writes: "[a]rguments, as I understand them, are expressed by a variety of sources. Most straightforwardly we have written or spoken texts that express arguments. I have no objection to saying that pictures, musical pieces, or even sculptures might express arguments. At the same time (…) pictures or musical pieces or sculptures will not themselves be arguments. In addition, while many texts that express arguments are themselves arguments, plenty are not, for any text with an implied conclusion is not itself an argument" (2003: 4). Conversational implicatures could fit nicely in this picture.

Ani carry out this interpretation and why is she justified in doing so since on the level of what is said the conveyed message differs from the intended one. In order for the belief to be justified the two components, namely what is said and what is meant have to be plausibly linked. In the interpretation of conversational implicatures there are two elements that carry a major practical and theoretical weight, namely, the context and the intention of the speaker. Conversational implicatures cannot exist in a vacuum, they need a specific context to be triggered. The same utterance will not give rise to an implicature in every context. In our example the utterance "Mix yourself a drink" conveys the message "It will take some time for me to get ready" only because it is a reaction to the question "How much longer will you be?" asked presumably in a context in which Ani and Ben are going somewhere together, let's say on a date. Additionally, there seems to be a requirement for the idea that there is a communicational intention behind the utterance that invites the hearer to infer a message that is different from what is said. This requirement is posed by the fact that we treat other people as rational if there is no evidence to the contrary.[3]

The notion of intention[4] is the starting point for the reconstruction of conversational implicatures. More specifically, according to Grice, conversational implicatures are always (potentially) calculable. This calculation is not a psychological process that occurs while the hearer interprets the utterance, but is a post facto reconstruction (see Haugh 2008) that competent speakers should be able to carry out according to the following scheme:

> [The speaker] has said that p; there is no reason to suppose that he is not observing the maxims, or at least the Cooperative Principle; he could not be doing this unless he thought that q; he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that q is required; he has done nothing to stop me thinking that q; he intends me to think, or is at least willing to allow me to think, that q; and so he has implicated that q (Grice 1989: 31).

[3] Here the notion of intentional stance would be of help, at least considered in relation to the prediction of human behaviour: "[h]ere is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in most instances yield a decision about what the agent ought to do; that is what you predict the agent will do" (Dennett 1989: 17).

The notion of "theory of mind"/"mindreading" is also unavoidable: "[t]hat humans are capable of mindreading is all too obvious. We attribute mental representations to one another all the time. We are often aware of what people around us think, and even of what they think we think. Such thoughts about the thoughts of others come to us quite naturally" (Mercier and Sperber 2017: 94).

[4] Taken as the ascription of intentions to the utterer by the audience (see Sbisà 2001).

This calculation is used as evidence for the argumentative rationality associated with conversationl implicatures (see Sbisà 2006 and 2007).

Let's see another example.
Ani: Are you ready for the movie?
Ben: I am tired.

Again, we have a dating context. Ani and Ben decided to go to the movies together but when the time came Ben said that he was tired. In this kind of situation, the natural thing to do for Ani is to interpret his utterance as cooperative and relevant, which is in accordance with what Grice calls the Cooperative Principle: "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged." (Grice 1989:26) Grice also individuates four maxims related to the Cooperative Principle—Quantity (make your contribution as informative as it is required), Quality (try to make your contribution one that is true), Relation (be relevant) and Manner (be perspicuous) (Grice 1989).

If Ani is a competent, and I will claim, a rational speaker, she will naturally be guided by the Cooperative Principle and the maxims and will be able to reconstruct the inferential process that let her to the interpretation of the speaker's message following the scheme mentioned above. The steps of her reasoning can be reconstructed as follows:

(1) I assume B is following the rule of relevance.
(2) His remark would not be relevant unless the fact that his fatigue is relevant to whether or not he is ready for the movie.
(3) I know that when people are tired they often do not wish to go to movie dates.
(4) If Ben is tired he does not wish to go to the movies.
(5) Ben probably assumes I will reason in this way, and has not said anything to stop me from doing so.
(6) I conclude that Ben intends to convey that he does not wish to go to the movies.

What I would like to do next is claim that somewhere inside this reconstruction lies an argument, or in other words, that conversational implicatures express arguments.[5]

3. *Conversational implicatures and arguments*

The connection between implicatures and arguments has been recognized before. Macagno and Walton, coming from a dialectical framework, claim that

> (…) conversational implicatures represent implicit meaning triggered by the use of a sentence and (…) they can be considered interpretations of

[5] We could also say that conversational implicatures are arguments, but this formulation should not be taken in a rigid literal sense.

> the meaning of a word or a speech act. On this perspective, conversational implicatures are triggered by conflicts of dialogical and epistemic presumptions that are resolved by a process of best explanation, which in turn are based on argumentation schemes such as inference to the best explanation, practical reasoning, argument from sign, appeal to pity and analogy. Depending on the context, the presumptions on which the process of explanation is based on vary, and therefore the conclusion of the implicit argument can be different. (2013: 223)

According to the authors, conversational implicatures are explanations for presumptive inconsistencies. We have seen in the two examples presented above that the hearer can be taken as starting her interpretation by wondering why the speaker has said exactly what he did and not something else, perhaps more informative and direct. The authors argue that conversational implicatures "need to be analyzed as implicit arguments, involving a pattern of reasoning leading from a specific premise to a conclusion" (Macagno and Walton 2013: 211).

According to the systematization presented by the authors our second example could be explained as an Argument from Cause. It is an instance of causal argumentation that links an event to its effects. The general scheme of argument from cause can be represented as follows:

"Major Premise: Generally, if A occurs, then B will (might) occur.
Minor Premise: In this case, A occurs (might occur).
Conclusion: Therefore in this case, B will (might) occur"
(Macagno and Walton 2013: 219).

In our second example, the speaker replies that he is tired instead of giving a direct negative answer to the question whether he is ready to see the movie. The goal of the utterance is not to inform the interlocutor about his fatigue, but to lead her to draw a conclusion from cause to effect. Tiredness is presumed to be incompatible with going to the movies: if someone is tired, he needs to stay at home, and if someone stays at home, he cannot at the same time see a movie at the theater. This causal relationship is presented as an alternative: either A or B; not A; therefore B. (Macagno and Walton 2013: adapted from 219).

At this point it is necessary to take a step back and see what kind of arguments could conversational implicatures possibly be. Trying to answer this question Moldovan (2012) states the following:

> [i]n order to avoid confusions it is relevant to point out that 'argument' is sometimes used to refer to a speech act of arguing, and sometimes used to refer to an abstract object, which is the content expressed by speech acts of arguing. (…) On the other hand, 'implicature', although sometimes used to refer to the content of an act of implicating something, it was introduced by Grice as a technical term to name the act of meaning that q by saying that p (…) To avoid confusion, the question under discussion here should be formulated as follows: is a conversational implicature always a speech act of arguing? And he concludes: "I think the answer to the above question should be negative. (2012: 304)

If we are faced with the choice between a speech act of arguing and an abstract object we could be tempted to deny that implicatures are in any way related to arguments. They are indirect speech acts so they cannot be considered merely as abstract objects[6] and since conversational implicatures are not characterized by a disagreement between the involved parties they cannot be considered as instances of the speech act of arguing. Consider again our examples. In both of them Ani asks a question to which Ben replies. He does not do so directly and Ani has to reach the intended message inferentially, but there is no disagreement that the parties need to solve. There is a communicative exchange aimed at the sharing of information. This sharing is achieved by the understanding of the implicature at play which can be represented as the reconstruction of an argument.

Luckily for us, the meaning of the words "argumentation" and "argument" are not exhausted by Moldovan's two options. At this point we can turn to a distinction proposed by Hitchcock in a general discussion about argumentation that will be useful in our current discussion about the connection between conversational implicatures and arguments: "[i]n English, the word 'argument' and the corresponding verb 'argue' are used in two quite clearly distinguishable senses." (Hitchcock 2017: 448). In the first sense

> arguing requires only one arguer (who in cases of collaboration in the production of an argument can be a group of people). The arguer expresses a point of view on a question, and offers as support for this position one or more reasons. The expression of the point of view and the provision of one or more reasons in its support constitute a complex of speech acts. The arguer addresses these speech acts to one or more readers, listeners or observers, who need not reply. (Hitchcock 2017: 448)

As we can see, according to the first sense arguments are produced by a speaker to put forward his point of view and the reasons he has to support it. The second sense presupposes two or more interlocutors sharing opinions.

> The other sense is that in which we say such things as "they were arguing with one another" or "they had a bitter argument" or "she argued with him". In this sense, arguing requires at least two arguers; if one argues with oneself in this sense, then one sequentially takes two different roles. The arguers express to each other divergent opinions on some question. Each attempts to get the other(s) to accept their point of view, not necessarily by offering reasons in support of it. (Hitchcock 2017: 449)

As we can see from the above quotes, according to Hitchcock an argument can provide supporting reasons for one's position or can be aimed at getting the other person to accept the arguer's position. He labels the first kind of argument the "reason-giving" sense and the second the "disputational" sense of "argument" and "argue". The application of this distinction to conversational implicatures is fairly simple. I would

[6] For a discussion about the ontological status of arguments see for example Goddu 2010, Sinard-Smith and Moldovan 2011 and Patterson 2013.

like to claim that conversational implicatures are reason-giving arguments in which the speaker (arguer) addresses a hearer who does not need to reply. In the case of conversational implicature the speaker is not trying to convince the hearer to accept a position she disagrees with but is explicitly stating a reason in support of his intended message.

As we have seen earlier explaining our second example as an Argument from Cause, if someone is tired he needs to stay at home, and if someone stays at home, he cannot at the same time see a movie. The explicit premise "I am tired" provides a reason for not going to the movies, that is, for the unstated conclusion. The general way of thinking of the form "if A occurs, then B will (might) occur" is the missing premise that links the explicitly given reason to the conclusion. Because of that we can reconstruct the argument on which the conversational implicature is based. The utterance is one of the premises, more precisely it is the only explicit part of the argument since all additional premises, as well as the conclusion, are unstated.

4. *Conversational implicatures and enthymemes*

Now we have arrived to the most unintuitive part of the view of conversational implicatures as arguments, or more precisely, utterances expressing arguments, namely the fact that in this strange kind of argument only one premise is explicitly stated, while the conclusion and one or more additional premises are implicit. The idea is that conversational implicatures can be considered a special case of enthymemes, that is, instances of arguments with unstated premises or conclusions. Enthymemes are reconstructed on the basis of their explicit elements using deductive, inductive or abductive forms of reasoning. These forms of reasoning differ in the level of strictness they possess. Induction is associated with statistical inference, deduction is not defeasible and abduction is characterized by plausible reasoning that can admit exceptions. The missing premises are generally taken to be assumptions that are needed to make the argument valid. The attribution of assumptions will often be justified by appealing to the principle of charity by which we should attempt to supply a missing statement that makes the argument valid or at least to choose the interpretation that makes the argument stronger. We have to keep in mind that the person we are attributing the conclusion to has never actually made that claim explicitly. Because of that we can say that enthymemes are not the same as the reconstructed arguments based on them. We can say that the reconstructed argument represents the original one. The same goes for conversational implicatures. The reconstructed argument is not the same as the utterance, but it can be a representation of it and of its underlying structure. The two are not the same, but are closely related. According to Gilbert (1991) incomplete arguments should be filled in with missing assumptions that are plausible to the intended audience or recipient of the argument and that appear to fit in with the position

advocated by the arguer, as far as the evidence of the text indicates. If we apply this idea to conversational implicatures taken as reason-giving arguments we could say that incomplete arguments should be filled in with missing assumptions that are plausible to the hearer and that appear to fit in with the position advocated by the speaker, as far as the evidence of the context indicates.[7]

If we accept that communication sometimes requires from us the interpretation of indirect arguments, it could be asked what are the advantages of such communicative strategy? Is there any additional reason why speakers would choose incomplete forms of argumentative reason giving besides communicational economy if there is always a degree of uncertainty that accompanies indirect communication? Jackson and Jacobs (1980) claim that enthymemes can be considered a special instance of Grice's Quantity Maxim: be as informative as necessary, but avoid being more informative that is necessary. They claim: "Enthymemes are not built the way they are for reasons of economy (i.e., merely to avoid the unnecessary); their method of construction optimally exploits the rules of turn taking so as to respect the preference for agreement. Giving too much support for an assertion or proposal is not merely pointless, but positively detrimental. Giving more support than is necessary increases the number of places where disagreement may occur—and does so without improving prospects for agreement." (264) Again, we should try to apply this idea to conversational implicatures as reason-giving arguments. If we accept that with the use of conversational implicatures the speaker provides to the hearer a reason to accept an implicit conclusion on the basis of that reason, qua explicit premise, and other unstated premises we could explain the appeal of conversational implicatures in the following way: in most cases the hearer will arrive, if she is cooperative and competent enough, to the intended message effortlessly and unconsciously. She will accept, or at least recognize as present, this unstated message that we have characterized as the conclusion of an argument. If she has reached this conclusion, we can assume that she has also individuated, or that she is capable to reconstruct, the inferential steps that lead to this conclusion, that is, that she can individuate the argument expressed by

[7] The arguments expressed by implicatures will always be characterized by a degree of uncertainty, but this is not specific for them: „[u]nlike verbal arithmetic, which uses words to pursue its own business according to its own rules, argumentation is not logical business borrowing verbal tools; it fits seamlessly in the fabric of ordinary verbal exchanges. In no way does it depart from usual expressive and interpretive linguistic practices. Statements with logical connectives (or other logical devices), and even sequences of such statements that more or less correspond to syllogisms, are just part of normal language use. They are used by speakers to convey a meaning that cannot be just decoded but that is intended to be pragmatically interpreted. Not only the words used but also the force with which premises and conclusions are being put forward are open to interpretation. They maybe intended as categorical or as tentative assertions, hedged by an implicit "in normal conditions." (Mercier and Sperber 2017: 163, 164).

the utterance. Since she has drawn the conclusion herself she cannot refute the whole inferential process directly. If she tried to refute the unstated conclusion the speaker could (truly!) say that he never said that. She could try to refute the stated premise but doing so in isolation from the whole argument would yield strange results—looking back at our examples, could Ani really say things like "You are not tired" or "I can't mix myself a drink" without Ben being confused?

It could be claimed that this picture puts a heavy burden on the hearer, leaving to the speaker the opportunity to distance himself from the indirect message, that is, to cancel the implicature. The idea that the hearer is somehow responsible, or even guilty for the believes he forms on the basis of indirect messages, is not uncommon in epistemology and ethics[8]. Still, I believe that this interpretation is erred. The view of conversational implicatures as argument gives more support to the hearer to justify her believes formed through indirect communication than it gives room to the speaker to cancel his message. Let's return once again to our examples.

Ani: Are you ready for the movie?
Ben: I am tired.

According to the standard interpretation, Ben is too tired to go to the movies and this is the message that he wishes to convey to Ani. She will reach the intended message taking into account Ben's cooperativeness, his communicative intention and the balance between the relevance of what is said and what could have been implicated. Still, this is not the only option. We can imagine a situation in which Ben wants to go to the movies and in which his comment about his tiredness is just a passing remark that carries no additional communicative weight. In this situation, the reasoning could be the following: he never said that he does not want to see the movie and Ani inferred that the movie date is cancelled on her own. She could easily have stopped at the level of what is said. We can also imagine that this interpretation could be used by Ben in a fight between him and Ani a couple of days after the date should've taken place. On that occasion, Ani could complain that they never go anywhere, mentioning the cancelled movie date as a recent example and Ben could claim that he never said that he did not want to go to the movies. This could be true, which opens up two possibilities: it could all be a misunderstanding or Ben could have been manipula-

[8] For example, in the domain of testimony, Fricker writes: "[t]hese acts all share with paradigm tellings the successful getting across of a message. I shall not investigate here the respects in which they differ; except to say that where what is conveyed is not explicitly asserted there is, I believe, a diminution in the responsibility for the truth of what is got across incurred by the utterer" (2006: 246–7). Writing about misleading (what we could call intended false implicatures) Adler notes: "[d]epending on the nature of the deception, the victim feels anything from foolish or tricked to corroded. Not only has he been misled, but the embarassment or horror of it is that he has been duped into collaborating on his own harm. Afterward, he cannot secure the relief of wholly locating blame externally" (1997: 442).

tive. Even if it is undeniably true that Ben did not actually say that he does not want to go to the cinema I would claim that Ani was correct in reaching this interpretation.

To sum up, there are, I suggest, three interpretative possibilities in this case. Ben either wanted to implicate that he does not want to go to the movies, to mislead Ani into thinking that, or he made a communicational mistake. In each scenario, the responsibility for the belief that he does not want to go to the movies formed by Ani on the basis on Ben's utterance can be justified taking into account the context of the utterance, the relevance of Ben's answer and the general communicational practice in which stopping at the level of what is said would be considered weird, uncooperative and even irrational. Think for example of the request "Can you pass the salt?" Stopping the interpretation of this utterance on the level of what is said and answering with "Yes, I can" without actually passing the salt to the utterer would be in most cases considered as a silly joke. The general idea is that competent language users should take conversational implicatures into consideration while communicating with other people; they should be aware that speakers could use them and that speakers could seek for indirect messages in what they said. If the context is known to both parties, in most cases, the interpretation should yield a true belief. Not only that, if the context is known to third parties they should also agree, at least to a certain degree, to one specific interpretation. This makes it possible for the reader to understand the examples presented in this paper, as well as in all other papers concerned with indirect communication. That is because the intention of the speaker is not important per se. What is important is the possibility to reasonably attribute an intention to the speaker and then build on it by relying on the best theoretical and normative considerations.

The possibility of the reconstruction of an implicature as an argument makes the justification of Ani's belief even more openly a process that is not private and unconceivable by anyone other than the person who formed the belief in question. In a particular context, for reasons of communicational relevance, the utterance of the speaker should be taken as a premise on the basis of which the hearer (and all other interested parties) could reconstruct an argument whose conclusion is the intended message. As we have seen it could be suggested that by adding the missing premises and accepting the conclusion of this implicit argument the hearer is somehow responsible for accepting the conclusion. Still, I suggest that accepting the implicated conclusion is the only cooperative and rational thing to do in a situation in which the hearer can rationally presuppose that the intention of the speaker is to send a message that is different from the uttered one. This presupposition is made plausible by the application of the notion to instrumental rationality to indirect speech. The guiding question is how to make sense of the utterance. If the direct reading does not satisfy the requirement of relevance, that is, it is not informative enough for the current

communicational exchange, under the presumption that the speaker is instrumentally rational we should look for a reasonable explanation of his behavior. I suggest that a reasonable explanation could be the idea that he intended to communicate something else—this makes indirect communication an instrumentally rational means to achieve a certain communicative goal. To find out what he wants to communicate we have to start from the presupposed intention and reach a conclusion by filling in the gaps, that is, providing the missing premises of the argument expressed by the implicature.

5. *Conversational implicatures as inferences or explanations*

One thing we need to look into is the possible objection that we should consider conversational implicatures as merely inferences or explanations. I would like to maintain the position that there is more to them and that it is best to consider them as pragmatic phenomena that express arguments.

Allan (2001) compares the inferencential process related to conversational implicature to the case in which someone concludes that the person on the other side of a telephone line is female because the speaker has a high-pitched voice. This inference could be incorrect and, as claimed by the author, so could those related to implicatures. The way implicatures generate belief and knowledge is probabilistic even if it is based on deductive reconstruction, but there is a difference between the inference present in the phone case and that related to the understanding of conversational implicatures that can bring them closer to arguments. As we have seen, conversational implicatures are a conversational endeavor just like the act of argumentation. On the other hand, the inference made on the basis of a high-pitched voice on the other side of the telephone is private, it is not meant to share information or create a belief in another person. Also, there is no premise in the phone case, at least not a verbal one. In a conversational implicature there is one explicit premise, which is the utterance intentionally addressed to the hearer. This is a direct invitation to reach the intended message, that is, the conclusion of the argument that can be reconstructed. I believe that this kind of communicative invitation marks a clear difference between conversational implicatures and inferences like the one present in the phone call example.

Talking about the difference between arguments and inferences, Mercier and Sperber write the following:

> [a]n inference is a process the output of which is a representation. An argument is a complex representation. Both an inference and an argument have what can be called a conclusion, but in the case of an inference, the conclusion is the output of the inference; in the case of an argument, the conclusion is a part—typically the last part—of the representation. The output of an inference can be called a "conclusion" because what characterizes an

> inferential process is that its output is justified by its input; the way however in which the input justifies the output is not represented in the output of an intuitive inference.
> What makes the conclusion of an argument a "conclusion" (rather than simply a proposition) is that the reasons for drawing this conclusion on the basis of the premises are (at least partially) spelled out (2011: 58).

As I wrote before, my suggestion is that the utterance that triggers the implicature can be seen as the only explicit premise in the argument that can be reconstructed on its basis. This premise carries the reason to accept a certain conclusion in a certain context. It is an invitation to make an inference and a commitment of the speaker to the relevance of what is said. Seeing the connection between an utterance and its conversational implicature as the link between a premise and a conclusion is in contrast with the stance that in the case of conversational implicature the speaker wants to convey a message only loosely related to what is said.[9]

To sum up, we can say that the difference between the reconstruction of implicatures and any other kind of reconstruction related to inferences is the following: the reconstruction of implicatures starts from an explicit premise—from "what is said"—and relies thus, at least partially, on other people, their input, their intentions, the common ground etc. In this context, "what is said" is an invitation for the hearer to draw an inference. This makes the creation and interpretation of implicatures an intrinsically communal endeavor. On the other hand, inferences can be private and often there is no linguistically coded input.

Now we can take a look at the relation between conversational implicatures and explanations. Discussing the relation between arguments and explanation Govier writes the following: „[a] fundamental difference between arguments and explanation is that in arguments, premises are intended to provide reasons to justify a conclusion whereas, by contrast, in explanations claims are put forward to show how a phenomenon came to be. In an explanation, someone tries to explain why some claim is true, whereas in an argument a person tries to demonstrate that it should be accepted" (2014: 14).

If we apply this reasoning to conversational implicatures seen as arguments, we can say that in an implicature the premises (stated and unstated) are reasons to accept the conclusion. The existence of an im-

[9] When Bach writes about the difference between conversational implicature and impliciture he notes the following: "Impliciture is to be distinguished from Grice's (1967a) conversational implicature. In implicature one says and communicates one thing and thereby communicates something else in addition. Impliciture, however, is a matter of saying something but communicating something else instead, something closely related to what is said." (1994: 126). This formulation can be considered misleading and it certainly goes against the intuitions presented in this paper. If conversational implicature is not closely related to what is said how can we ever reach the intended message? Seeing this connection as the link between a premise and a conclusion could be one way to explain the relation between what is said and what is implicated.

plicature can be explained by reconstructing the argument beneath it, but this will include additional elements, such as a general theoretical notion of implicatures themselves. We could say that the argument that can be expressed by a conversational implicature can be part of the explanation of the existence of the implicature, but the explanation will have to include nations such as context, intention, relevance and so on.

I conclude by pointing out the usefulness of an approach that treats conversational implicatures as speech phenomena that can be said to express arguments, or that can be reconstructed as arguments. The general idea is that if the use and interpretation of conversational implicatures are guided by argumentative reasoning this phenomenon can have a normative aspect: we could ask if there is a right way to interpret them or if we can talk about communicational responsibility for what is implicated. I would like to propose a normative view of the use and interpretation of conversational implicatures that can be applied in epistemology and ethics. Communicational responsibility[10] for what is said can be supported by the view of implicatures as arguments since if the speaker is presenting a premise to the hearer he is inviting her to reach a conclusion. If the hearer is competent and rational she will do so. The burden of proof in case the speaker wants to distance himself from the conclusion lies on him and not on the hearer.[11] He has to prove that the hearer has reasoned wrongly. In this paper, I wanted to argue for the idea that it will be more difficult to do so if we consider conversational implicatures as speech phenomena expressing arguments. The underlying argument gives strength to the conversational implicature and to the belief formed by the hearer on its basis.

I have used different formulations to describe the relation between arguments and conversational implicatures: in a loose sense it could be said that they are argument, we can say that they can be reconstructed as arguments, and we can talk about conversational implicatures expressing arguments. The best wording could perhaps be found by ask-

[10] The notion of "communicational responsibility" could be linked to the notion of "epistemic culpability", a term from social epistemology which describes a "failure to respond to evidence in the appropriate way" (see for example Begby 2013). The notion of "epistemic culpability" is present primarily in the discussion about epistemic injustice and prejudice and is primarily linked to the "hearer". I use the notion of "communicational responsibility" primarily in relation to indirect communication, but it could be considered in a much broader sense as the commitment of a competent language user to the belief his interlocutor forms on the basis of his or her words, assuming that the hearer is also a competent language user in the sense that he or she respects general linguistic, interpretative and rational standard during the belief forming process.

[11] Of course, there could be cases in which the speaker ha san implicature in mind but the hearer does not reach the intended conclusion or rejects the conclusion and takes into consideration only what is conveyed directly. In such cases, we should consider the hearer as not competent enough or as not cooperative. The argumentative view of implicatures could be also used in these cases to give strength to the implicated message.

ing who produces the argument—is it created by the speaker or by the hearer? I believe that it is best to associate the argumentative aspect of conversational implicatures with their reconstruction and to make it part of the justification of a hearer's belief. Nevertheless, the first input (the explicit premise) in the argument has to be provided by the speaker in the form of an utterance. As the reconstruction of conversational implicatures in general, the reconstruction of implictures as arguments is only potential. It does not have to be actually carried out while interpreting and understanding an utterance but it is important that competent and rational language users could provide such reconstruction. Even if only potential, this reconstruction is crucial since it creates the argument that justifies the belief held by the hearer. The reconstruction and the argument can be presented to the speaker as a justification of the belief held by the speaker and as proof of his communicative responsibility.

6. *Conclusion*

This paper should be considered as one of the first steps towards a more comprehensive theory of communicative responsibility. Such theory is needed since it is often claimed that the hearer is, at least partially, responsible for a belief she formed on the basis of implicated content she attributed to the speaker even if this attribution is justifiable.

I want to suggest that if a rational and competent hearer formed her belief taking into account all the factors needed for the understanding of a conversational implicature, she should be considered a cooperative and competent language user. Even in the case the belief she formed turned out to be false due to malicious intentions on part of the speaker or because of a misunderstanding, it is wrong to consider her even partially responsible for holding false beliefs since forming the belief in question was, from a communicative, and subsequently an epistemic and moral point of view, the right thing to do in that particular context. I believe that this position could be reinforced by considering conversational implicatures as speech phenomena that express arguments. In this way we do not use pragmatics to explain and advance our knowledge of argumentation but vice versa, we use argumentation to give strength to indirect communication.

The idea that conversational implicatures express arguments could be unintuitive, after all, meaning q with p could be just that, an inference from what is said to what is implicated. Still, I would like to claim that conversational implicatures are not merely inferences or explanations, as could also be suggested. This is due to the fact that a competent language user should always be capable of reconstructing the process that led him to conclude q from p. This reconstruction will have the form of an argument in which there is only one stated premise. On the basis of that premise and a reasonable attribution of intention to the speaker the hearer will fill in the gaps of the argument in question

relying on the context, general knowledge of the used language and other relevant factors.

We can consider conversational implicatures as reason-giving arguments in which the speaker (arguer) addresses a hearer who does not need to reply. In those cases, the speaker is not trying to convince the hearer to accept his position, as in the case of disputational arguments, but is explicitly stating a reason in support of his intended message.

By grasping the intended message, the hearer intuitively accepts the reasons provided for it and can also reconstruct the argumentative path that lead from an explicit reason, qua premise, to the intended conclusion. This could be interpreted as a justification for the view that the hearer is responsible for the belief she formed on the basis of conversational implicature, but I would like to claim that it actually justifies the idea that forming her belief as she did was the only cooperative and rational thing to do.

References

Adler, J. E. 1997. "Lying, deceiving, or falsely implicating." *Journal of Philosophy* 94 (9): 435–452.

Allan, K. 2001. *Natural Language Semantics*. Oxford: Blackwell.

Bach, K. 1994. "Conversational Impliciture." *Mind and Language* 9 (2): 124–162.

Begby, E. 2013. "The Epistemology of Prejudice." *Thought: A Journal of Philosophy* 2 (2): 90–99.

Dennett, D. 1989. *The Intentional Stance*. Cambridge: MIT Press.

Fricker, E. 2006. "Testimony and Epistemic Autonomy." In J. Lackey and E. Sosa (eds.). *The Epistemology of Testimony*. Oxford: Clarendon Press: 225–250.

Gilbert, M. A. 1991. "The Enthymeme Buster: A Heuristic Procedure for Position Exploration in Dialogic Dispute." *Informal Logic* 13 (3): 159–166.

Goddu, G. C. 2003. "The Context of an Argument. OSSA Conference Archive. 28" https://scholar.uwindsor.ca/ossaarchive/OSSA5/papersandcommentaries/28.

Goddu, G. C. 2010. "Is 'Argument' Subject to the Product/Process Ambiguity?" *Proceedings of Proceedings of the 2010 ISSA Conference*. Amsterdam. June 29.

Govier, T. 2014. *A Practical Study of Argument*. Wadsworth.

Grice, P. 1989. "Logic and Conversation." In P. Grice. *Studies in the Way of Words*. Cambridge: Harvard University Press: 22–40.

Haugh, M. 2008. "The place of intention in the interactional achievement of implicature." In J. L. Mey and I. Kecskés (eds.). *Intention, Common Ground and the Egocentric Speaker-Hearer*. New York: De Gruyter.

Hitchcock, D. 2017. *On Reasoning and Argument: Essays in Informal Logic and on Critical Thinking*. Berlin/New York: Springer.

Jackson, S. and Jacobs, S. 1980. "Structure of conversational argument: Pragmatic bases for the enthymeme." *Quarterly Journal of Speech* 66 (3): 251–265. DOI: 10.1080/00335638009383524.

Macagno, F. and Walton, D. 2013. "Implicatures as Forms of Argument." In A. Capone et al. (eds.). *Perspectives on Pragmatics and Philosophy*. Berlin/New York: Springer: 203–224.

Mercier, H. and Sperber, D. 2011. "Why Do Humans Reason? Arguments for an Argumentative Theory." *Behavioral and Brain Sciences* 34 (2): 57–74.

Mercier, H. and Sperber, D. 2017. *The Enigma of Reason*. Cambridge: Harvard University Press.

Moldovan, A. 2012. "Arguments, Implicatures and Argumentative Implicatures." In H. Jales Ribeiro (ed.). *Inside Arguments: Logic and The Study of Argumentation*. Cambridge: Cambridge Scholars Publishers.

Patterson, S. W. 2013. "Are arguments abstract objects?" *OSSA Conference Proceedings*. Ontario Society for the Study of Argumentation.

Sbisà, M. 2001. "Intentions from the other side." In G. Cosenza (ed.). *Paul Grice's Heritage*. Brepols: Turnhout.

Sbisà, M. 2006. "Two Conceptions of Rationality in Grice's Theory of Implicature." In E. Baccarini and S. Prijić-Samaržija (eds.). *Rationality in Belief and Action*. Rijeka: Filozofski fakultet.

Sbisà, M. 2007. "On argumentative rationality." *Anthropology & Philosophy* 8 (1/2): 89–100.

Simard Smith, P. L. and Moldovan, A. 2011. "Arguments as Abstract Objects." *Informal Logic* 31 (3): 230–261.

Croatian Journal of Philosophy
Vol. XVIII, No. 53, 2018

On the Rationality of Conspiracy Theories

DANIEL COHNITZ
Utrecht University, Utrecht, The Netherlands

Conspiracy theories seem to play an increasing role in public political discourse. This development is problematic for a variety of reasons, most importantly because widespread belief in conspiracy theories will undermine the institutions of open societies. One of the central questions that will need to be answered here if we hope to find out why conspirational thought is recently gaining such support and to find out how to respond to it, is the following: what mindset leads to the belief in conspiracy theories? People who believe in conspiracy theories are often ridiculed as nutcases, tinfoil hats, and paranoid crackpots, while they portray themselves as particularly critical, better informed and enlightened responsible citizens. Finding out which of these characterizations is correct is crucial for coming up with the appropriate response to the rise of conspirational thought. In this article, I want to discuss this question and the phenomenon of conspirational thought in two respects. First, I want to explain how philosophy, and epistemology in particular, is essential for understanding the phenomenon and for developing a strategy to deal with the harmful kind of conspirational thought. Secondly, I want to show how epistemology in turn can learn from studying this phenomenon.

Keywords: Social epistemology, conspiracies theories, normativity of rationality, epistemic authority, conspiracy theory of society.

1. *Introduction*

It surely feels as if conspiracy theories and conspirational thought are at an unprecedented high. A quick glance into the comment section of almost any topic at almost any online news outlet will acquaint you with a wide range of conspiracy theories, from the moderately suspicious to the outright bizarre.

Indeed, it seems that almost any recent event is the result of a conspiracy. 9/11 was an inside job, Obama's birth certificate is a forgery,

global warming is a major hoax. Malaysia Airlines Flight 370 has been abducted by the Unites States, Malaysia Airlines Flight 17 has been shot down by a Ukrainian fighter jet in a failed attempt to assassinate Vladimir Putin. Refugees are let in masses to enter the EU as part of a large conspiracy that tries to abolish nation states and install the New World Order. The Clintons are members of a child sex ring that operates from a Washington D.C. pizza place. The Paris terror strikes of 2015 were in fact another inside job, while the terror attack on a Christmas Market in Berlin last winter did not actually kill anyone—with the exception of the alleged terrorist—but was just entirely staged by the German government.[1]

For all we know, these theories are false. More importantly, some of them are simply absurd. Can't we then just shrug the rise of conspiracy theories off as an insignificant—and perhaps even somewhat entertaining—cultural development that is largely due to the fact that—thanks to the internet—nerdy crackpots have it now very easy when they want to disseminate their crazy views?

Well, I don't think that we can just shrug it off. On the one hand, conspiracy theories, of the kind just listed, aren't harmless. As I said, for all we know, they are false. So, people who believe these theories believe something false, and might—because of that—make poor choices. They might end up voting for the wrong party, or oppose important policies, they might not vaccinate their children, might not be willing to contribute to efforts intended to prevent global warming, and so forth.

And these are still only the moderate negative consequences. On December 4 of 2016, 28-year-old Edgar Maddison Welch took his assault rifle and a revolver and drove from Salisbury to the pizza restaurant "Comet Pingpong" in Washington D.C. in order to investigate himself whether the restaurant indeed houses a sex ring of child abusers (as it was claimed in the "pizzagate" conspiracy theory). He did fire his gun there, fortunately not hurting anyone, and, of course, found nothing. So, one reason for being worried about the rise of conspirational thought should be that it leads to false beliefs.

False conspiracy theories are also harmful to those not believing them. Scientists, officials and journalists will have to spend their valuable time debunking false allegations. This debunking is enormously complicated by the fact that the false theory one is trying to debunk is a conspiracy theory. All the evidence one can produce against the conspiracy theory is just too easily interpreted in the light of that very theory as simply being further smokescreen produced by the conspirators.

Finally, conspiracy theories are not only a symptom of a receding trust in expertise, science, and the government, but the propagation of these theories further nourishes this distrust. This development

[1] All claims about the content of particular conspiracy theories in this essay can be verified via the relevant Wikipedia entries. These entries also contain links to websites providing more information.

undermines these institutions. In a society in which universities and experts aren't trusted anyway, few will protest when universities are closed and research funding is cut.

So, these theories aren't harmless. But their support isn't marginal either, it isn't confined to "middle-aged white male Internet enthusiasts who live in their mother's basements" (Uscinski and Parent 2014: 5). As polls in the United States show, "conspiracy theories permeate all parts of American society and cut across gender, age, race, income, political affiliation, educational level, and occupational status" (Uscinski and Parent 2014: 5).

The political relevance of this phenomenon is also apparent in Europe. 'Lügenpresse' (*lying press*) has become famous as the fighting word of the self-proclaimed "critical citizens" that support PEGIDA or the AfD in Germany. Essentially, the background of the "Lügenpresse" allegation is just a conspiracy theory, according to which the left and liberal mainstream media are collaborating with the German government in its evil plan to destroy Germany with an influx of immigrants.

This all is surely reason enough to take the apparent rise of conspirational thought seriously and to inquire into its origin. Many social scientists, psychologists, historians, and a somewhat smaller group of philosophers have thus looked into that issue in the past few years.

One of the central questions that will need to be answered here if we hope to find out why conspirational thought is recently gaining such support *and* to find out how to respond to it, is the following: *what mindset leads to the belief in conspiracy theories?* People who belief in conspiracy theories are often ridiculed as nutcases, tinfoil hats, and paranoid crackpots, while they portray themselves as particularly critical, better informed and enlightened responsible citizens. Finding out which of these characterizations is correct is crucial for coming up with the appropriate response to the rise of conspirational thought. Is the best response logic and argumentation or is it therapy and medication?

In this essay I want to discuss this question and the phenomenon of conspirational thought in two respects. First, I want to explain how philosophy, and epistemology in particular, is essential for understanding the phenomenon and for developing a strategy to deal with the harmful kind of conspirational thought. Secondly, I want to show how epistemology in turn can learn from studying this phenomenon. Along the way, I hope it becomes clear that I see the analysis of conspiracy theories, its current popularity and its cure a matter that requires interdisciplinary effort.

2. *What Can Philosophy Contribute?*

So let's begin with the question how philosophy can contribute. *That* philosophy can contribute might not be obvious. Sociology and psychology might *prima facie* be seen as the most relevant disciplines to explain the phenomenon of the rise of conspirational thought and the

questions of whether its proponents are nutcases or sane critical citizens.

2.1. *The Sociology of Conspiracy Theories*

Let us consider sociology first. Conspiracies themselves are a social phenomenon. Under which conditions they can form, be sustainable and successful is an issue that social scientists are best placed to investigate. It is also a matter of sociology to inquire who (i.e. what type) develops theories about conspiracies and how these theories get disseminated to wider audiences and come to be influential for public discourse. The latter is the social study of—what might be called—"conspiracy culture" (Aupers 2012).

Some sociologists believe that the study of such cultures should abstain from any normative judgments about the rationality or accuracy of the beliefs that are held and sustained in these cultures. In the literature on conspiracy theories I found this view motivated from either Max Weber's considerations for a value-free science, or by David Bloor's conception of the so-called strong programme (e.g., Harambam (2017) cites both motivations as a motivation to refrain from any normative judgments). But also sociologists who do *not* shy away from taking a normative stance are not terribly helpful in identifying what if anything is wrong with conspiracy theorists. Joseph Uscinski and Joseph Parent, for example, in fact try to identify standards by which one is supposed to measure how likely it is that a given conspiracy theory is true (Uscinski and Parent 2014). However, these considerations then play no role whatsoever in their sociological analysis of how widespread conspiracy beliefs are in the American public.

Clearly, the social sciences can tell us a good deal about certain *descriptive* matters. Like, for example, under what conditions conspiracy theories are likely to spread, and what type of people typically ends up believing these. The great value of the analysis by Uscinski and Parent, for example, lies in the observation that conspirational thought in the US, although also there it seems to many to be on the rise, is actually over the past decades in decline.

They also found, as already reported above, that belief in conspiracies is not confined to specific demographics in the US. Moreover, they found that the public impact of a conspiracy theory very much depends on who's in the White House. Under democratic presidents, more people believe that communists are the conspirators, under republican presidents more people believe that it's capitalists who conspire against the American public. Of course, in 2014 Uscinski and Parent couldn't foresee that in 2017 a president would be in the White House who himself propagates conspiracy theories.

Another sociological find, that seems to be quite robust, is the fact that conspiracy theories often lead to more conspiracy theories. That is, if someone starts to believe that 9/11 was an inside job, such person

will often also believe in other conspiracies, such as the idea that vaccinations cause autism, or that climate change is a hoax or that the holocaust didn't happen. Moreover, it seems that there doesn't need to be any coherence between these theories, other than that they explain events in terms of a conspiracy (Sutton and Douglas 2014).

As we will see below, such observations are crucial in order to understand the rise of conspirational thought in a society, but by itself they don't answer the question to what degree (if any) conspiracy theories are *rationally* believed. For that latter question, the social sciences do not provide the necessary expertise.

2.2. *The Psychology of Conspiracy Theories*

One might perhaps think that psychology is better equipped to deal with that latter problem. Psychology informs us about our unconscious biases and the quick and dirty heuristic that our mind uses to generate solutions to problems that would sometimes better be tackled by careful reflection. Psychology also informs us about character traits and personality profiles, all of which can then be correlated with certain types of beliefs. For example, a prominent explanation for why some people might be prone to believe in conspiracy theories is that they have a hypersensitive module for agency detection.

Hypersensitive agency detection is considered a cognitive bias that came about as an adaptation.

> [S]ince humans have evolved in an environment that contains many agents (e.g., friends, enemies and dangerous predatory animals) hypersensitivity to agency may be adaptive because it makes people wary in their interactions with the environment around them, reducing vulnerability to unexpected outcomes and avoiding risk from potentially dangerous factors. Being able to detect and understand an event and react quickly, or respond quickly to an ambiguous situation, is important for physical and social survival. (Douglas et al. 2016: 60)

While taking a stroll through the park, you hear a noise in a nearby bush or tree, and you spontaneously form the belief that someone or some animal is hiding there, while perhaps the sound just came from the wind stirring up some leafs. Hypersensitive agency detection means that you sometimes suspect agency when in fact there is none. Having a reaction of that kind might still have been overall better for survival even if it sometimes leads to mistaken assumptions of agents in your environment. As so often, it's better to be safe than sorry. The fact that we have this hypersensitivity to agency has been cited to explain why humans believe in the existence of invisible spirits and gods, why we are superstitious and belief in causal connections among unconnected events, etc.

Obviously, to overascribe agency and intentionality and to see purposes and causal connections where there are none is typical for false conspiracy theories. Thus, plausibly, people that tend to overascribe

agency might also be prone to believe in conspiracies. As Douglas and colleagues have shown in their (Douglas et al. 2016) this is indeed the case. They found that a low level of education predicts endorsement of conspiracy theories, which is mediated by a general tendency to over-attribute intentional agency. Thus, some people end up believing in conspiracies because of a certain thinking style that involves hypersensitivity to agency.

Again, this is a very valuable and interesting result, but whether this is going to help in understanding how we should respond to the apparent rise in conspirational thought is still a question that is itself left open by these psychological results. Is it sometimes rational to believe in a conspiracy theory? If it is not rational to believe in such theory then the existence of a hypersensitive module for agency detection could explain why some people nevertheless believe in such theories. But the psychological result is silent on that question. We learn that low educational level predicts belief in conspiracies. But whether you believe a theory *rationally* is often a matter of your background knowledge, a matter of the evidence that you have for the theory and the alternative explanations that you are aware of. Perhaps people who lack certain levels of education lack the background knowledge that would make their belief irrational. In order to understand better whether it is at all possible to believe rationally in a conspiracy theory, we should eventually turn to philosophy.

2.3. *The Philosophy of Conspiracy Theories*

Theoretical philosophy deals in several areas with the kind of normative questions we are here seeking answers to. In *logic* we investigate the logical correctness of reasoning and argumentation, in *epistemology* theoretical philosophers investigate the conditions under which we can have knowledge of the world around us, and in *philosophy of science*, we look at the questions of when theories are supported by evidence, what criteria an explanation needs to satisfy in order to be a good scientific explanation, and how scientists should go about testing and revising their theoretical accounts of the world. This is precisely the kind of expertise that seems relevant for answering the question when—if ever—it might be rational to believe a conspiracy theory. So let us see what theoretical philosophy has to say about conspiracy theories.

The philosophical engagement with conspiracy theories is still relatively young and—unfortunately—as yet not very developed. It is often (e.g. Pigden (1995), Coady (2012)) claimed that Karl Popper was the first to write about it in his famous opus magnum *The Open Society and its Enemies* in 1945 (Popper 1945).

Popper's theory is supposed to be that conspiracy theorists must believe that every event is due to intentional successful planning and Popper holds against this view that it overlooks the fact that many if not most of the consequences of our actions are not in fact under our

control. Very often events happen as unintended consequences of our actions. Hence conspiracy theories are irrational, because they rest on an untenable assumption concerning the amount by which we have control over the consequences of our actions. If many or most of these consequences are not intended, then it can't be true that all events are the product of successful intentional planning.

People who understand Popper as making this argument have been quick to point out that Popper's critique of conspiracy theories can't be right. Why should every conspiracy theorist assume that *all* events are the result of successful intentional planning? After all, conspiracy theories are typically theories of specific outcomes, for example the collapse of the World Trade Center, or the death of Lady Di. Why should conspiracy theorists then have to believe that *all* events that happen were so planned? And, clearly, *sometimes* events *do* come out as planned, so how is Popper's argument supposed to work?

I believe that this controversy is due to a misunderstanding of Popper's writings. Popper is not in fact engaging with conspiracy theories in the way in which we are interested in them here. Popper is instead trying to make a quite valid methodological point about the social sciences in general. When he talks about the "conspiracy theory of society" he has sociologists in mind who think that the social sciences work by providing intentional explanations of social events.

A theory of the type Popper has in mind here, is—what he calls—*Vulgar Marxism*, the idea that social events are to be explained by identifying the social class-related motives of the protagonists that brought an event about, and to explain the event, in turn, as the intended satisfaction of these class-related motives. But that's not a theory or a criticism of conspiracy theories as such.

So Popper didn't believe that all conspiracy theories rest on that one mistaken premise that *all* conspiracy theorists must believe that *all* events are due to a conspiracy. Nevertheless, this is a "type" of view that one can indeed find in the philosophical literature. There are several attempts to show that belief in conspiracy theories rests on some fundamental mistake, such that it is always or almost always irrational to believe such a theory, just because it is a conspiracy theory.

This is certainly somewhat in line with the ordinary use of the word 'conspiracy theory'. To call a theory a "conspiracy theory", or someone a "conspiracy theorist" is often intended in a derogatory way. Wondering whether conspiracy theories are irrational—in that usage—makes no sense; of course, they are. This usage is so widely spread that conspiracy theorists themselves want to avoid the label. For example, you find YouTube videos on one of the paradigmatic conspiracy theories, the chemtrails theory, that are titled "Chemtrails are not a conspiracy theory".

If that's so, then perhaps we should define 'conspiracy theory' right away as a certain type of irrational belief. So, perhaps *a conspiracy*

theory is an irrationally believed theory that explains an event as the result of a conspiracy. The problem with such definition is that it doesn't relate to the psychological and sociological research on conspiracy theories. Most of that research tries to correlate belief in certain specific explanations—for example the theory that 9/11 was an inside job, or that Oswald didn't act alone—with a demographic or psychological profile. But this correlation will hold between this profile and conspiracy theories as defined in the proposed definition only if these theories are irrationally believed by the people that participate in that study. But these studies typically do not investigate on what evidence or with which justification these theories are believed (as we have seen above in the discussion of whether a hypersensitive agency detection module can *explain* belief in conspiracy theories).

The trick to close this gap is not to define conspiracy theory outright as an irrationally believed explanation of some event, but to show that conspiracy theories *can* only be believed in an irrational way.

It seems that what we'd need in order to draw immediate consequences from the empirical results of psychology and sociology is a critical analysis of conspiracy theories that resembles in result David Hume's critical analysis of miracles. In section X of his *Enquiry* David Hume makes the following argument:

> A miracle is a violation of the laws of nature; and as a firm and unalterable experience has established these laws, the proof against a miracle, from the very nature of the fact, is as entire as any argument from experience can possibly be imagined. (Hume 1748)

Hume defines a miracle as a violation of the laws of nature. Thus, to *recognize* something as a miracle is to recognize it as a violation of the laws of nature. But that means that we must have had strong empirical evidence to believe the law in the first place, which the miracle supposedly violates. Which means in turn, so Hume's reasoning, that whatever evidence we think we have for the proposed miraculous event is simply outweighed by the evidence that speaks against it—it is far more likely that we are not dealing with a miracle but with a mistaken or misleading observation report. Consequently, believing that a miracle has occurred is always irrational. This analysis rests on a conceptual and an empirical component. The conceptual component is the definition of a miracle as a violation of the laws of nature. The empirical component of the argument is human fallibility, and the fact that miracles are never observed by sufficiently many in order to outweigh the evidence against them.

Hume's argument is controversial, but my point is independent of whether Hume is right about miracles. I just want to explain by that analogy that if we had such an analysis for conspiracy theories as Hume offers for miracles, we could immediately answer our initial question: conspiracy theorists must be paranoid nutcases, because there is no rational way to believe such a theory (just as—according to Hume—there

is no rational way to believe in miracles). Also, sociological and psychological research into the correlation between, say, cognitive biases and conspiracy belief, would then directly tell us something about the causes of these beliefs.

Alas, I don't quite see how such a Humean analysis could be achieved, at least not on conceptual considerations alone. Such "conceptual considerations" of course depend on the definition one presupposes for the term 'conspiracy theory'. We have seen above that such a definition should—at least for our purposes—not include *explicitly* that any believe in such theory is irrational.

But what is a conspiracy theory then? First of all, it is an explanation for sometimes an event, sometimes just some other kind of phenomenon. For example, the inside job conspiracy theory of 9/11 explains as an event the collapse of the World Trade Center, and the theory that Diana, Princess of Wales, was killed by MI6 explains as an event the car crash in a tunnel in Paris in August 1997. As in these cases, the events in question are often tragic or even traumatic, and believing that they were the results of malicious intentional planning might be a form of psychological coping with the tragedy.

However, not every conspiracy theory explains a traumatic event. For example, the chemtrails conspiracy theory holds that long-lasting trails, so-called "chemtrails", are left in the sky by high-flying aircraft and that they consist of chemical or biological agents deliberately sprayed for purposes undisclosed to the general public. So, this theory explains the durability of the condensation trails of planes.

In the introduction, I already mentioned the pizzagate conspiracy theory. This theory developed quickly after WikiLeaks released emails that were hacked from the account of John Podesta, the chairman of Hilary Clinton's campaign in the presidential elections of 2016. The publication of these emails occurred just a month before the election. The emails where then discussed on social media and a few Trump supporting platforms. At some point, it was suggested that the occasional reference to pizza and to meetings at Washington restaurants were in fact secret code for child pornography (allegedly, 'cheese pizza' with the initials 'c' and 'p' is code for 'child pornography'). Under this new interpretation, seemingly innocent and unconnected emails about dinner invites were quickly revealed to be in fact hiding a child sex ring that the Clintons and the Podesta brothers were a central part of, and that operated from the restaurant "Comet Pingpong" in Washington D.C. In the case of the pizzagate conspiracy theory, there is no specific event or even tangible phenomenon that gets explained. Instead, the theory explains why John Podesta exchanged emails that contained reference to pizza and dinners at restaurants.

So far, we know that conspiracy theories are explanations of a certain kind that explain something. What makes these explanations now *conspiracy* theories is that they invoke secretly conspiring agents as a

salient cause in the explanation. The reasons why these agents are conspiring *secretly* can be several. Often it is assumed that the conspirators are not up for anything good. They have malicious, evil plans, that would meet with heavy resistance if the public were to learn about them.

But the maliciousness of the intentions behind the conspiracy are not a necessary ingredient of conspiracy theories, even if one restricts the analysis to the paradigmatic cases. First of all, some conspiracy theories remain agnostic about the intentions that are behind the conspiracy. For example, a prominent German authority on the chemtrails conspiracy theory expresses in interviews that he does not know what exactly the chemtrails contain and what they are good for, and that it is conceivable that they are supposed to serve some beneficial purpose.

A clearer example is perhaps the theory that Paul McCartney died in a car crash already in November 1966. According to that theory, Paul drove off after an argument with the other band members during a recording session, crashed his car and died. He was then replaced by a certain "William Campbell" who had previously won a Paul McCartney lookalike contest. William Campbell is in conspiracy circles referred to as "Faul McCartney".

The theory is extremely elaborate and there are hundreds of clues found that are supposed to support it. At first, the theory is supposed to explain why the Beatles for some time after Paul's death did not appear in public together. Since Faul McCartney is supposed to be taller than Paul, they also didn't play many live concerts anymore with the new line-up (so people wouldn't notice the difference in height between Paul and Faul). But the band also left clues for their fans, since they couldn't quite keep that secret for themselves. Most prominently, the cover of the Beatle's *Abbey Road* record, which displays, according to the conspiracy theorists, a funeral procession. Lennon in white is the priest, Starr symbolizes the undertaker, Harrison, in denim is the gravedigger, and Faul, finally, out of step with the others and barefooted is the supposed.

The theory is also confirmed by several backside messages that the Beatles allegedly hid in their songs, and it explains the complex symbolism of the cover of the *Sgt. Pepper's Lonely Hearts Club Band* (1967), which supposedly represents the funeral of Paul McCartney.

According to most versions of that theory that I came across, the band, the management and the media together conspired to hide the death of Paul McCartney from the public. But the intentions behind it were not primarily sinister. Most theorists seem to hold that the death of McCartney was covered up to spare the public from grief. The Beatles were so popular at the time that it would have been a catastrophe for the fans to learn that their idol had died.

Also, according to the conspiracy theory, the cover-up had an altogether positive side effect; Faul is supposedly more talented than Paul McCartney and the music of the Beatles improved after Paul's death.

Thus, sinister motives are often assumed to be the reason for the agents acting in secrecy, but it isn't a defining feature of a conspiracy theory.

Some definitions of conspiracy theories have a problem with the secrecy condition. It is clear that conspirators *intend* to act in secrecy. But can a theory still be a conspiracy theory if the secret conspiracy has been leaked by, for example, a whistleblower? Once something is out in the open such that others are aware of it, it isn't anymore a secret, but isn't being secret a defining feature of conspiracy theories?[2]

Again, this seems to be a confusion. Otherwise everyone who sincerely proposes a conspiracy theory would thereby undermine it. Conspiracy theories couldn't be rationally believed, because they—literally—couldn't be believed. Obviously, the secrecy requirement just means that the conspirators *intend* to act and coordinate secretly.

Another feature that has been suggested as a defining feature of conspiracy theories is the fact that they are often in conflict with the "official story". On that view, the official explanation of the collapse of 9/11, namely that it is due to a conspiracy between members of the terror network Al-Kaida, is not a conspiracy theory, while the theory that the collapse is due to a conspiracy in the US government is a conspiracy theory.

There are two problems with this requirement that a conspiracy theory always has to be in conflict with the official story. One is that what counts as "the official story" is context dependent. Russian mass media came up with several explanations of the crash of Malaysian Airlines Flight 17 that conflicted with the explanation that the Dutch Safety Board and the Dutch-led Joint Investigation Team provided. According to the story that was in Russia considered to be the official account, the Ukrainian army was to blame, while the official story recognized in the Netherlands holds that the plane was shot down by a missile that was fired from a rebel-controlled area. Which of these theories is a conspiracy theory and which is not doesn't depend on whether you are in Russia or in the Netherlands.

Another, related problem is that a conspiracy theory doesn't cease to be such if its proponents come to power. On April 10, 2010, an aircraft of the Polish Air force with several Polish government officials on board crashed near the city of Smolensk in Russia. For some time, the official story according to Polish and Russian investigators was that the plane crashed because of an unsafe landing approach in the bad weather conditions of that day. In 2015 the political party "Law and Justice" won the Polish parliamentary elections; the leader of that party, Jaroslaw Kaczyński believes that the crash is due to an assassination—possibly orchestrated by Russia—and the investigation of the crash has since been reopened. What used to be a conspiracy theory in conflict with the official account is ince 2015 a story supported by officials.

[2] See Räikkä (2009) and Mandik (2007).

Likewise, the fact that in 2017 it is the current president of the United States, Donald Trump, who publishes on Twitter that his predecessor, Barack Obama ordered to wiretap his phones during the election campaign, doesn't make the wiretapping story any less of a conspiracy theory. So, again, *being in conflict with the official story* is not a defining element for a conspiracy theory.

But then we arrive at a rather thin definition of the term 'conspiracy theory':

Definition. A *conspiracy theory* is an explanation that cites agents acting together in secrecy as a salient cause.

This definition is thin, which means that it is also quite broad.[3] The official explanation that 9/11 was the result of a secret plot by Al-Quaeda terrorists is a conspiracy theory just as much as the theory that it was an inside job. But since we left out sinister motives from the definition, also your suspicion that your friends may be planning a surprise birthday party for you, is a conspiracy theory.

But because the definition is that broad, we also know immediately that conspiracy theories in this sense can be rationally believed. In fact, all of us believe several conspiracy theories. In other words, a general argument that could show that conspiracy theories are always irrational to believe and that would just fall out of a definition of 'conspiracy theory' is not forthcoming. The kind of argument that Hume produced against miracles cannot be produced against conspiracy theories.

3. *What is wrong with conspiracy theories?*

But that means that a lot more work needs to be done, both on the side of philosophy and on the side of sociology and psychology to understand what goes wrong with conspiracy theories *when* they go wrong.

Now, of course, often conspiracy theories will be bad explanations of an event for familiar reasons. Familiar in the sense that philosophers of science have already identified such reasons. For example, one explanation of an event might be worse than an alternative explanation of that same event if it can explain less aspects of the event, if it is less supported by the empirical evidence gathered about the event, if it needs to postulate a greater number of unlikely events, etc.

Also, theorists who fall for these inferior explanations might make familiar mistakes, they might have taken a too uncritical attitude towards the available evidence, they were epistemically lazy in not reviewing alternative explanations, and they might have been ignorant with respect to certain parts of the available evidence.

Psychological research into a "conspiracy mindset", as well as philosophical research into epistemological virtues and vices, can then profile typical conspiracy theory believers and characterize their typical mistakes.

[3] A similarly broad definition is defended in Dentith (2014).

My suspicion is, though, that these familiar pitfalls of bad reasoning and bad epistemological practice will only partly explain the phenomenon of the rise of conspirational thought, at least for Western Europe. And I believe that this is also the area in which philosophy might gain new insights from studying this phenomenon.

In societies in which most of the media is clearly partisan, and where your main or only channels of information are one-sided and also unreliable, you may be lazy and end up with a largely mistaken picture of the world around you. The echo-chamber that will reinforce your mistaken beliefs is set up for you, and when you don't make an effort to break out of it, you will remain having a world view that might be massively mistaken.

Fortunately, the relatively open societies of Western Europe are not like this. For the most part, the media are doing a pretty good job in providing accurate and relatively balanced information, they are also relatively independent, and from the way they are organized and managed, not likely to get under external influence that could force the media to distort the information they are providing. In such a situation, being an epistemically lazy citizen of such a society does not automatically lead to inaccurate beliefs. You actually need to do something, you need to find alternative information sources in order to become massively misinformed.

And, strangely enough, this is what happens. The conspiracy theories that *you* find in the comments section of your favourite news-outlet have others found in the internet on specially dedicated blogs and websites, when looking exactly for this alternative account. The people writing these comments do not lazily believe what the mainstream media tell them, but go the extra mile and look up "information" that we safely ignore. After critically weighing the different accounts they then make up their own mind, and—tragically—end up with spectacularly false beliefs.

It is true that thanks to social media, like *Facebook*, people can now create their own echo chambers in which they only receive the information that supports their world-view and beliefs, but not all conspiracy theorists confine their information intake to such echo chambers. Indeed, those who post the weird comments under *your* favourite news outlet, must at least also consult that source. The picture of the lazy, gullible, ignorant conspiracy theorists (c.f. (Cassam 2016)) seems inappropriate for many cases.

This corresponds to the image that conspiracy theorists have of themselves. They are *sceptics*, they look at the information they receive via mainstream channels more critically than others and are enlightened and better informed than the average citizen (see Harambam (2017)). Indeed, the strategies they use in choosing the theory to believe are often consistent with recommended criteria for good explanations (which I already listed above), choose the theory that can explain more

aspects, choose a theory that is supported by more evidence, choose a theory that doesn't postulate a great number of unlikely events.[4] Of course, in the application of these criteria, mistakes are made, mistakes that are due to the fact that laypersons are often not in a good position to apply the appropriate criteria for theory choice.

I believe philosophy and epistemology in particular can gain from a better understanding of these mistakes, because it promises to lead to a better understanding of the general epistemological principles which philosophy tries to formulate.

Some of the mistakes are relatively easy to identify. Take for example the standard *cui bono* heuristic that conspiracy theorists use in order to find the conspiracy that caused the event they try to explain. Sometimes such a heuristic makes sense. Finding the culprit of, say, a murder by asking who would have had a motive, is a useful strategy if you know that the event in question was murder; an intentional and planned killing. But you can't just use this heuristic randomly for any event that you want to explain, since not every event is the result of successful intentional action. Here Popper's argument against the conspiracy theory of society, that we encountered earlier in this paper becomes now relevant for ordinary conspiracy theories. Using a *cui bono* heuristic without prior independent evidence that an event was indeed caused by intentional action relies on an unreasonable assumption about the world and is thus irrational.

Other mistakes are not that easy to analyse. Conspiracy theorists often misidentify the relevant experts. They mistrust the proper scientist, but put trust in charlatans. But how should they have known who the proper expert is, given that they themselves don't know the relevant subject matter? What criteria can laymen use to determine whom they can trust? In social epistemology, this is discussed as the "Novice/2-Experts Problem", a problem that will require a solution in terms of indirect indicators of relevant expertise (see Goldman (2001)). Such indicators that will have to be provided by trustworthy institutions, such as universities, but these institutions are in the danger of losing public trust, and we need to understand why that is and what we can do to stop and reverse this development.

Finally, conspiracy theorists seem to suffer from a certain overconfidence in their own ability to inform themselves and arrive at a considered judgment over issues for which they don't possess the relevant expertise. Again, it is difficult to blame them for this if epistemologists standardly recommend that the art of critical thinking requires the exercise of your own informed, critical judgment. Apparently, there is something wrong with the idea that we *always* should exercise *our own* best judgment. Sometimes it seems just prudent to trust the experts and to defer to their epistemic authority. When that's so and whether

[4] The observation that conspiracy theories often seem to satisfy criteria for theory choice better than their competitors is also made in Hepfer (2015).

such deference is compatible with the enlightenment ideal of epistemic autonomy is an open question that philosophers need to answer (see Zagzebski 2013)).

I hope that I have shown that conspiracy theories provide a fruitful test-case for philosophical theories and that philosophy is the relevant discipline to provide the normative analysis of the rise of conspirational thought in Western societies. How the normative analysis should then be translated into policies and strategies to address this dangerous phenomenon is, however, a question that philosophers, sociologists, psychologists and political scientists will have to solve together.

References

Aupers, S. 2012. "'Trust no one': Modernization, paranoia and conspiracy culture." *European Journal of Communication* 27: 22–34.

Cassam, Q. 2016. "Vice epistemology." *The Monist* 99: 159–180.

Coady, D. 2012. *What to believe now: Applying epistemology to contemporary issues*. Oxford: Wiley-Blackwell.

Dentith, M. R. X. 2014. *The Philosophy of Conspiracy Theories*. London: Palgrave Macmillan.

Douglas, K. M., et al. 2016. "Someone is pulling the strings: hypersensitive agency detection and belief in conspiracy theories." *Thinking & Reasoning* 22: 57–77.

Goldman, A. 2001. "Experts: Which ones should you trust?" *Philosophy and Phenomenological Research* 63: 85–110.

Harambam, J. 2017. *The truth is out there: Conspiracy culture in an age of epistemic instability*. PhD thesis. Erasmus University Rotterdam.

Hepfer, K. 2015. *Verschwörungstheorien: Eine philosophische Kritik der Unvernunft*. transcript Verlag.

Hume, D. 1999. *An Enquiry Concerning Human Understanding*. Oxford: Oxford University Press.

Mandik, P. 2007. Shit Happens. *Episteme* 4: 205–218.

Pigden, C. 1995. "Popper revisited, or what is wrong with conspiracy theories?" *Philosophy of the Social Sciences* 25: 3–34.

Popper, K. R. 1962. *The Open Society and Its Enemies*, volume II, The High Tide of Prophecy: Hegel, Marx and the Aftermath. London: Routledge Kegan Paul.

Räikkä, J. 2006. "On Political Conspiracy Theories." *Journal of Political Philosophy* 17: 185–201.

Sutton, R. M. and Douglas, K. M. 2014. "Examining the monological nature of conspiracy theories." In J. W. van Prooijen and P. A. M. van Lange (eds.). *Power, politics, and paranoia: Why people are suspicious of their leaders*. Cambridge: Cambridge University Press: 254–272.

Uscinski, J. and Parent, J. (2014). *American Conspiracy Theories*. Oxford: Oxford University Press.

Zagzebski, L. 2013. "A defense of epistemic authority." *Res Philosophica* 90: 293–306.

Croatian Journal of Philosophy
Vol. XVIII, No. 53, 2018

Wisdom and Reason

ANDREI MĂRĂȘOIU
University of Virginia, Charlottesville, USA

On Ryan's (2012) theory of wisdom as deep rationality, to believe or act wisely is to believe or act in a justified way, informed by a body of other justified beliefs about the good life. Ryan (2017) elaborates the view along evidentialist lines: one's belief or act is justified when it is based on the best available evidence. The resulting package faces counterexamples. Transformative experiences are rational 'leaps of faith' (Paul 2014), so the agent's decision to undergo one is not best supported by the evidence available. Many transformative experiences (such as deciding to become a mother, or choosing a career path) often endow lives with meaning, and agents with a sense of purpose (Wolf 2010). Because so much is at stake, it is sometimes rational for agents to take on the risk involved in transforming themselves. Deciding to undergo such experiences may be wise—even if the evidence available at the time doesn't positively support that decision. In reply to this challenge, I argue that, instead of evidentialism, Ryan's view should include virtue theory, which helps explain the seeming counterexamples. I focus on the virtues of openness to experience, and of steadfastness in the face of experience.

Keywords: Wisdom, rationality, evidence, virtues, transformative experiences.

1. *Introduction*

What is wisdom? Sharon Ryan (2012) has advanced an influential theory on which wisdom is deep rationality. According to her view, to believe or act wisely is to believe or act in a justified way, and informed by a body of other justified beliefs about moral, emotional, and practical affairs. (These are beliefs needed to live well.) As Ryan (2017) elaborates, to believe or act justifiedly, in turn, is to believe or act based on the best available evidence.

Here is a reason in favor of Ryan's view. The view is intuitively compelling because it is procedural. It contrasts with success views, according to which being wise presupposes knowledge of how to live well,

and enacting that knowledge in actually living well. Success views of wisdom fail to accommodate the widely recognized fact that sages can find themselves in adverse circumstances, and wisdom may be a "burdened virtue" (Tessman 2005): a virtue that shines through precisely when things go wrong. When things go awry, we may still say someone acted (or thought) wisely in case they have nothing to reproach themselves. That is, they acted (or thought) in a justified way, a way that withstands reflective scrutiny.

In this text, I raise, and then respond to, an important objection to Ryan's view. Ryan (2017) advocates, as part of her view of wisdom, an account of rationality as justification, and an evidentialist account of justification. The resulting package, I argue, faces important counterexamples. The counterexamples are some transformative experiences. As Laurie Paul (2014) conceives them, transformative experiences are rational 'leaps of faith'. If so, then the agent's decision to undergo them is *not* best supported by the evidence available in that decision situation, contra Ryan's evidentialism. Many transformative experiences (such as deciding to become a mother, or choosing a career path) often endow lives with meaning, and agents with a sense of purpose (Wolf 2010). A lot is at stake in deciding to transform oneself. Because so much is at stake, it is sometimes rational for agents to take on the risk involved in deciding to transform themselves. Deciding to undergo such experiences may be wise—even if the evidence available to the agent at the time doesn't positively support that decision.

I respond to this challenge. Ryan's view of wisdom as deep rationality shouldn't be abandoned. But its evidentialist component should be renounced. A better companion for Ryan's deep rationality view of wisdom, plus her justification view of rationality, is a virtue theory of thinking and acting well. This, I argue, can help explain the seeming counterexamples to Ryan's view coming from transformative experiences. Suppose we take acts and thoughts to be rational inasmuch as they exercise skills and virtues their agent possesses. (This is how Zagzebski (1996) characterizes justification.) Then deciding to undergo transformative experiences—or not—may be rational in many circumstances: whenever one appropriately, and carefully, exercises one's virtues of openness to experience, or, respectively, steadfastness in the face of experience. Virtue theory comes to the rescue of Ryan's view of wisdom as deep rationality.

2. *Wisdom as deep rationality*

In a nutshell, here is Sharon Ryan's theory of wisdom as deep rationality:

> A person S is wise iff (1) S has a wide variety of epistemically justified beliefs on a wide variety of subjects that are central to a good liberal arts education, as well as morality, practical matters, and matters of the heart, (2) S has very few unjustified beliefs and S is sensitive to his or her limitations,

> and (3) S is deeply committed to learning more about the topics noted in (1) and living a life that reflects what S is justified in believing (2017: 117).

I will articulate, defend, and then respond to an important objection to Ryan's view in what follows. But I start by saying what makes her view intuitively compelling. Addressing both support and objections will clarify some of the central concepts of Ryan's view: what it is to be rational, what commitment consists it, and how wise thoughts and wise deeds cohere.

Ryan's view of wisdom as deep rationality is well-supported. It does justice to the fact that wisdom is most striking in adverse circumstances when, even if the sage has nothing to reproach herself, the outcome is not ideal. And Ryan's view also does justice to why the sage's thoughts and deeds cohere (to the extent that they do—which is also the extent of one's wisdom).

The core idea Ryan proposes is that of deep rationality. When is rationality *deep?* Ryan explains that it:

> requires that one be deeply committed to seeking out new ideas, becoming more educated, and testing one's own theories against all of the best evidence available... Furthermore, it requires the wise person to put his or her justified beliefs into practice. The theory requires the wise to be deeply committed to having appropriate emotions, treating others morally, and having successful strategies for getting through the trials and tribulations of life, etc. (2012: 109)

Ryan's conception of deep rationality depends on her view of rationality. When are thoughts or deeds *rational* in the first place? As she clarifies:

> I am going to focus on one, purely epistemic, sense of rationality. Epistemic rationality should be understood in terms of epistemic justification. I endorse an evidentialist theory of epistemic justification. (2017: 116–17)

That is, beliefs are rational when justified. Acts, presumably, are rational when the decisions to perform them are justified. To be justified is to be best supported by evidence (against countervailing evidence, if any there be).

Ryan (2017) lays great emphasis on evidentialism. However, I will now argue that this evidentialist construal of justification faces counterexamples: decisions which are not best supported by evidence, but which we would ordinarily often deem wise.

3. *Transformative experiences*

Consider what Laurie Paul (2014) calls *transformative* experiences: becoming a mother, choosing a career path, undergoing surgery to install a cochlear implant or a sensory substitution system so that one may perceive the world differently, undergoing a mystical experience of religious conversion, or abruptly quitting an addiction. Sometimes, undergoing a transformative experience changes one's life for the better, endowing it with meaning, and the experiencer with a sense of pur-

pose unsuspected before. In those cases, deciding to transform oneself is wise, or so we would ordinarily think. This is despite the fact that such decisions cannot be best supported by evidence in their favor. As Paul (2014) explains, they often amount to rational leaps of faith.

I will now argue that Ryan's view of wisdom, coupled with her evidentialism, faces a large class of counterexamples. If wisdom is deep rationality, and if rationality is justification, and if justification is evidentialist, then wise thoughts and deeds are, necessarily, those which are best supported by evidence. Deciding to undergo transformative experiences is sometimes a counterexample to Ryan's view because we often ordinarily deem such decisions to be justified even if they're not based on the best available evidence.

In what follows, I'll unpack this objection, giving the example of deciding to become a mother. If Ryan were right, wisely deciding to become a mother would involve deciding it based on the preponderance of evidence available to the person expecting. Does that always happen when the decision is wise?

As a preliminary to answering this question, consider what kind of evidence would be needed. For the purposes of both Ryan's deep rationality theory of wisdom, and Paul's conception of transformative experiences, evidence should be understood in an internalist way. The sage manifests deep rationality *in foro interno* because she is *committed* to being rational: this is an overarching project of hers. Not only should one's beliefs be justified; one's commitment to their justification should make the reasoner able to come with justifications *herself* if prompted. Not only should one's thoughts and deeds cohere; one's commitment to deep rationality entails that the reasoner be able to make that coherence *transparent* to herself if the question arises. Evidence relevant to justifying one's beliefs or decisions is evidence available internally for the agent's first-personal assessment.[1]

Does someone who is expecting always have this kind of evidence to support her decision? In cases of wise transformative decisions, the answer is no. As Paul (2014) describes it, undergoing a transformative experience changes our preferences in a way that often cannot be anticipated. One seldom knows beforehand—before giving birth—what one's preferences will be afterwards. When in that position, one cannot reason one's way through, starting from preferences prior to giving birth, and ending with one's preferences once one has a child. Suppose, for instance, that becoming a mother makes the child come first even if you were a party-goer and not much of a family person before. Often enough, you can't justifiedly anticipate whether this will happen or not at the time when you decide whether you will keep the pregnancy.

[1] It isn't enough for evidence *to exist* in favor of holding one's beliefs or making the decisions one makes in order for one to be wise. Rather, the sage has to be *sensible* to that evidence, having some dispositions to accurately detect and weigh evidence. In addition to the reason for preferring a virtue-theoretical approach to wisdom given in Section 5, this line of thought also supports it.

Nonetheless, whenever it does occur, the change in preferences involved in motherhood is a central part of what one might call a fulfilling or flourishing life. Plausibly, the life of a sage may include the joys parenthood brings. The experience of becoming a mother might constitute what it is for a woman to lead a good life, and her becoming wiser than before. Leading to wisdom, the decision itself would be a wise one. But that decision would fail to be supported by evidence.

4. *Evidence, high stakes, and risk*

One might reply to this kind of case by insisting that the decision to keep the pregnancy is based on evidence, only not on *conclusive* evidence. The idea is that, when you decide to keep the baby, you look around, see other mothers and other women who decided to terminate their pregnancies, read some relevant literature, and decide for yourself, weighing the pros and cons. True, the evidence you would be gathering is third-personal, and may not apply to *you.* This is why the evidence is fallible—but it is evidence still. Your decision—to the extent it is a wise one—is still based on evidence.[2]

This reply is in the right direction. Here is what is right about it. Often, a pregnant woman has to make this hard choice on scant evidence: third-personal evidence that some mothers are happy and others aren't, and first-personal evidence about her previous behavior, more or less suited for societal stereotypes of parenthood. The would-be mother has evidence both in favor of, and against, deciding to become a mother. How, in these circumstances, would it be rational for her to act on evidence that doesn't settle the matter? And, if not rational, her decision can't be wise either—at least if the deep rationality theory of wisdom is correct.

While in the right direction, the reply ignores a crucial fact. Deciding to become a mother often endows the parent's life with new and deeper meaning, and gives the parent a sense of purpose. (Often, not always.) *A lot* is at stake in deciding to become a mother. So her decision to keep or terminate the pregnancy is a decision with *high stakes.* The most common construal of rational acts is as acts which are performed considering what is at stake. So it is rational to commensurate what is at stake with what *risks* the agent can rationally undertake. To wit, if a lot is at stake, it is rationally permissible (albeit not compulsory) to risk somewhat to get what you want.

Contrast this with evidentialism about how to justify your decisions. A decision is justified only if the evidence, on balance, supports it. This contrasts sharply with a decision taken by assuming a risk afforded by

[2] It is important to see that the evidence relevant here is that available to the agent at the time of making the decision to undergo a transformative experience the consequences of which cannot be fully appreciated. What one learns after undergoing that experience cannot apply, in hindsight, to decide the difficult question of which decision should be made.

the high stakes being decided on. When you risk to achieve something you deeply desire—say, the meaningfulness of parenthood—you may well do so even when the evidence available to you, on balance, doesn't support that conclusion.[3]

To sum up, you may decide to become a mother—and that decision may be wise—even when the evidence does not support that decision, over and above the countervailing evidence. This line of thought doesn't necessarily undermine evidentialism in general. It does undermine evidentialism as an accurate construal of what makes deeds and thoughts which underlie *wisdom* rational. If Ryan's view of wisdom as deep rationality is to be preserved, it has to give up this tenet. What can it replace it with? I will suggest that virtue theory is a good replacement as a background theory of rationality.

5. *Openness and steadfastness*

I will now argue that transformative experiences are no counterexample to Ryan's view of wisdom as deep rationality—*as long as* one holds a virtue-theoretical view of what rationality is. The result (e.g., Zagzebski 1996) is a picture on which overt acts and mental acts (e.g., coming to believe or to desire something new) can be given a *unified* account in point of what makes them rational. Any such performance (overt or mental) is rational inasmuch as it exercises a skill or virtue the agent has. If what is at stake is practical rationality, the relevant skills will be practical and the relevant virtues will be moral (courage, generosity, etc.). If what is at stake is theoretical rationality, the relevant skills will be cognitive and the relevant virtues will be intellectual (intellectual humility, open-mindedness, etc.) If what is at stake is rationality *tout court*, all such skills and virtues will be relevant.

In this more encompassing sense of rationality, deciding to undergo transformative experiences may indeed be rational. Which skills or virtues would be manifest in transformative experiences, to make them rational? Their rationality would be accounted for by the intellectual virtue of *openness to experience.* Openness unhinged may lead to excess, hence to vice—a far cry from what reason would recommend. But openness to experiences for which we have good though defeasible third-personal evidence that they are significant for achieving a fulfilling life, that kind of openness allows us to see deciding to undergo transformative experiences as rational.

[3] Note that a friend of evidentialism might instead give up a tenet central to Ryan's deep rationality theory of wisdom, namely, that the rationality of beliefs and the rationality of decisions is to be evaluated in the same way. Evidentialists may insists their view concerns beliefs alone, not decisions too. Traditional evidentialists (e.g., Feldman and Conee 1985) may be construed this way. This reply would be to the point because the counterexamples given above all concern transformative decisions, not transformative episodes of coming to believe something new.

While a view that relates wisdom and deep rationality has to accommodate the wisdom of deciding to undergo transformative experiences, it has to also accommodate their exceptional status. And it can do so. Part of what Ryan means when she characterizes *deep* rationality is that one's commitment to self-improvement is deep—in having rational thoughts and doing rational deeds. What does such a commitment amount to? I think that such a commitment is a manifestation of steadfastness, the virtue of not wavering in front of adverse circumstances or luring experiences.

When should one be open to new experiences, and when should one hold steadfast in resisting them?[4] The answer is uniform. We should be open to new experiences of which we have good though defeasible third-personal evidence that they are (even ever so mildly) significant for achieving a fulfilling life. We should, however, hold steadfast when confronted with the prospect of insignificant experiences—experiences which don't enhance the meaningfulness of our lives. We should hold steadfast for a simple reason: when transformative, experiences are risky (who knows who or what I will then become?). And, if they don't enhance the meaningfulness of our lives, the risk comes at no gain and should not be taken.[5]

It goes without saying that one is open to meaningful transformative experiences, or steadfast in the face of insignificant seemingly transformative experiences, given what one believes. The sage's intellectual virtues, by leading her to believe what's right, support her practical virtues in doing what's right. And, in particular, they support the virtues of making the best decision about whether to undergo a transformative experience or not—the virtues of openness and steadfastness.

6. *Conclusion*

I have clarified and amended what I take to be the most promising view about the nature of wisdom, that advanced by Sharon Ryan (2012). My proposal puts together two ideas I find individually promising, and mutually reinforcing: virtue theory and Ryan's theory of wisdom as deep rationality. My version of Ryan's view accounts for what makes transformative experiences *rational,* when indeed they *are* so: exercising the virtues of openness and steadfastness appropriately. It also accounts for what commitment is. Commitment is steadfastness in holding on to the patterns of thinking and acting that are worth holding on to—in light of one's first—and third-personal evidence.[6]

[4] In delineating these options, I don't mean to deny the obvious—that one could postpone difficult choices whenever appropriate.

[5] The idea of significance to one's life (Wolf 2010) is, of course, relative to many factors, and the decision to undergo a transformative experience ultimately depends on context. But the core idea that the sage should lead a meaningful life seems quite intuitive in full generality.

[6] Writing this paper was made possible by financial support from the Jefferson Scholars Foundation through a John S. Lillard fellowship. I am grateful to an

References

Feldman, R. and Conee, E. 1985. "Evidentialism." *Philosophical Studies* 48: 15–34.

Paul, L. A. 2014. *Transformative Experience*. Oxford: Oxford University Press.

Ryan, S. 2012. "Wisdom, Knowledge and Rationality." *Acta Analytica* 27: 99–112.

Ryan, S. 2017. "A Deeper Defense of the Deep Rationality Theory of Wisdom: A Reply to Fileva and Tresan." *Acta Analytica* 32: 115–123.

Tessman, L. 2005. "Conclusion." In: *Burdened Virtues*. Oxford: Oxford University Press: 159–168.

Wolf, S. 2010. *Meaning in Life and Why It Matters*. Princeton: Princeton University Press.

Zagzebski, L. 1996. *Virtues of the Mind*. Cambridge: Harvard University Press.

anonymous reviewer for the journal *Acta Analytica* for very insightful questions that led to the current manuscript.

Croatian Journal of Philosophy
Vol. XVIII, No. 53, 2018

Book Reviews

Hugo Mercier and Dan Sperber, The Enigma of Reason*, Cambridge: Harvard University Press, 2017, 396 pp.*

In their new book *The Enigma of Reason* Hugo Mercier and Dan Sperber try to solve the philosophical problem made explicit in the book's title. The enigma, as Mercier and Sperber see it, is in the following: humans are animals, but there is certainly something that separates them from the rest of the animal kingdom. The *differentia specifica* for humans is our ability (capacity) to reason. Now the question arises how well do we reason. Most experimental evidence, from the Four card problem (Wason selection task) to the Linda problem (a form of the conjunction fallacy), points to the conclusion that we reason poorly. If reasoning is our *evolutionary superpower*, as Mercier and Sperber's note, then it is very hard to see why we are endowed with a flawed superpower. Why is it that we are so bad at something that is, presumably, our evolutionary advantage? And how has something we are so bad at evolved in the first place? These two questions sum up the enigma from the book's title.[1]

Firstly, we shall look at the structure and methodology of the book in question. We shall see how the chapters and subchapters are structured and how Mercier and Sperber's argumentation emerges form that structure. Also, we are going to examine their writing style and how it impacts their argumentation. Secondly, we are going to evaluate the main argumentative line of the book. We shall see on which basis does their argumentation stand and how do they defend it with specific arguments. Thirdly, we need to recognize possible problems regarding the book and finally give a closing remark.

The Enigma of Reason has five major chapters (excluding the Introduction and the Conclusion): *Shaking dogma, Understanding inference*, *Rethinking reason*, *What reason can and cannot do* and *Reason in the wild*. The chapters follow each other exceptionally well. Mercier and Sperber start with identifying the problem (*Shaking dogma*) then they transition into building up the ground floor for their argumentation (*Understanding inference*) after which they build up their case (*Rethinking reason*, *What reason can and cannot*) and in the end they try to preventively react to possible criticism (*Reason in the wild*).

The writing style is lighthearted and engaging and there can even be found humorous elements in the book. One of those moments is that in

[1] Mercier and Sperber are calling it "double enigma".

which the authors talk about how people use reasons to justify and explain their action to others. In this specific case, a man named Theodore Wafer has killed a victim of a car-crash, Renisha McBride, on his door step, because he mistakenly believed he was being attacked.[2] In Mercier and Sperber's words:

> Wafer, for instance, could, perhaps not with sufficient good sense but not absurdly either, see his reasons as objectively justifying his shooting the person at the door. Had he, on the other hand, given as his reason for having fired a shot that Elvis Presley was dead and that therefore life was meaningless, we would see this not as a genuine reason-based explanation, not even a defective one, but as an admission (or a claim) of temporary insanity. (112)

Humorous excursions like this one can be found throughout the book. Other than entertaining the reader these excursions serve another purpose. They help the reader to transition more efficiently from one dense line of argumentation to another. On a macro level, the book is engaging and lighthearted because of the way in which almost every major chapter begins, namely, with some kind of narrative exposition. Form Descartes' internal struggles about Reason and doubt over Sherlock Holmes' not so great deductive powers to the tragedy of Alphonse Bertillon,[3] these narrative expositions serve as an introduction to a certain philosophical problem that Mercier and Sperber are trying to solve. That being said, it should be noted that the argumentation of the authors does not rely in any way on these expositions. Their argumentation is made on independent grounds.[4]

Mercier and Sperber's argumentation is clear and precise. On top of that, they do not presume the reader's prior knowledge about the philosophical problems that they are engaging with. The authors define and explain every concept and theory that they discuss. This ranges from rudimentary concepts in logic (like *antecedents* and *consequences*) to the explanation of dual process theory. As a result, the reader can successfully follow a line of argumentation that goes from simple to complex.

Here is the content of *The Enigma of Reason*. In the first place we need to establish what this book is about and then we will go deeper into the authors' argumentation. In *The Enigma of Reason* Mercier and Sperber are trying to answer two questions:

1. How do we reason?
2. Why do we reason the way we do?

Now we need to position the subject at hand in a broader philosophical context. Mercier and Sperber are tackling with one of the core problems in philosophy today (even in the history of philosophy for that matter)—the nature of human reasoning. One does not venture on a journey like that overnight. *The Enigma of Reason* is a product of years of research, both empirical and theoretical. Throughout the years the authors have been building up their arguments in papers like *Why do humans reason? Arguments*

[2] Mercier and Sperber use a real-life example here.

[3] The real-life tragedy was actually about Captain Albert Dreyfus in what is historically known as "*The Dreyfus Affair*", but the tragedy of reason is about Alphonse Bertillon.

[4] What those grounds are shall be discussed later in this review.

for an argumentative theory, *Intuitive and reflective inference* and *Epistemic vigilance*.[5] In that regard they certainly seem to be totally competent for the challenge at hand.

In the first chapter, *Shaking dogma*, Mercier and Sperber try to identify the problem. Humans do engage in the act of reasoning but it seems we are extremely bad at it. In the history of philosophy there were two schools of thought on that matter. One school was trying desperately to argue that humans do in a way reason properly (to some extent at least), while the other gave up completely on the notion that humans can reason properly and concluded that we are irrational beings. Mercier and Sperber are formulating this historical exposition in the form of a trial (*Reason on trial*), which is very engaging and entertaining. The second part of *Shaking dogma* is dedicated to the contemporary analysis of human reasoning mostly based on empirical evidence. Here the authors argue that even contemporary theories like dual process theory cannot explain how people reason in the light of evidence that comes from research in the psychology of reasoning. They conclude that we need a new approach to human reasoning.

In the second chapter, *Understanding inference*, Mercier and Sperber are setting up the stage. Throughout this chapter they are defining the relevant concepts and notions that they will use later in their argumentation. Firstly, they are differentiating between two concepts: inference and reasoning. Inference is an *extraction of new information from information already available, whatever the process*. Reasoning is *the particular process of pursuing that goal by attending to reasons*. As we can see, reasoning is just one form of inference. The main point being the emphasis on reasons, which will come in play later. Secondly, and arguably the most important definition in this book, is the definition of intuitions. Mercier and Sperber define intuitions as *judgments or decisions that are justified without knowledge of the reasons that justifies them*. That being said, Mercier and Sperber are also claiming that intuitions are *a distinctive "metacognitive" category*. An example that the authors provide is the following:

> When you infer from your friend Molly's facial expression that she is upset, you are more or less confident that you are right. Your own cognitive states are the object of a "metacognitive" evaluation, which may take the form either of a mere metacognitive feeling or, in some cases, of an articulated thought about your own thinking. (65–66)

Thirdly, the authors use the concept of modularity in order to explain how humans infer (on a mechanical level). The biological modules can be defined as *autonomous mechanisms with a history, a function, and procedures appropriate to this function*. The concept of modularity (in this sense) was first introduced by the philosopher Jerry Fodor in his book *The Modularity of Mind*. The authors are building up on that notion with the concept of massive modularity. Lastly, they define metarepresentations. Metarepresentations are representations about representations (higher order representations). Mercier and Sperber argue that we rarely engage in this kind of thinking (except for philosophers who have a professional deformation of this kind). On the other hand, we do engage in one specific form of meta-

[5] There are other papers and publications that have contributed to the creation of The Enigma of reason, but these are the most notable ones.

representational thinking all the time. That form is called *mindreading*. Mindreading is a metarepresentational module that informs us about what others believe through our intuitions. A simple example by the authors is the following:

> When the woman in the waiting room looks at her watch and sighs, you guess not only that she believes that the doctor is keeping her waiting but also that indeed the doctor is keeping her waiting. From this, you may draw further inferences of your own, such as about how well the doctor keeps her appointments. (103–104)

In the third chapter, *Rethinking reason*, Mercier and Sperber are arguing their case. They begin with establishing what reasons are. According to them, reasons are social constructs. *They are constructed by distorting and simplifying our understanding of mental states and of their causal role and by injecting into it a strong dose of normativity* and they are primarily for social consumption. Reasons themselves are inferred. How are they inferred? By intuitive inference. And how does intuitive inference works? Well, according to the authors, inferences are made possible by the existence of regularities in the world. For example, I can walk outside and see that the pavement is wet, but I certainly cannot see that the pavement being wet is a reason to believe that it has been raining. But what I can do is intuitively infer the reason in question. We are using reasons in two distinct forms: retrospectively and prospectively. When we use reasons retrospectively we do it to justify and explain our actions or beliefs. When we use reasons prospectively we do it for individual or communicative purposes. When we use reasons individually we are engaging in an act of inquiry and when we use reasons communicatively we engage in an act of argumentation. The authors are mostly focused on justification and argumentation because they believe that to be the primary functions of reasons. This is also what the reason module is design to do. In order to better explain how intuitive inference works here is a simple example from *The Enigma*:

> You arrive at the party and are pleased to see that your friend Molly is there too. She seems, however, to be upset. When you have a chance to talk to her, you say, "You seem to be upset tonight." She replies, "I am not upset. Why do you say that?" Just as you had intuited that she was upset, you now intuit reasons for your initial intuition. Here are what your two intuitions

might be:

> First order intuition: Molly is upset.
> Metarepresentational intuition about your reasons for your first order intuition: the fact that Molly isn't smiling and that her voice is strained is what gives me reasons to believe that she is upset. (136–137)

In this example we see how intuitive inference works. Intuitive inference operates on two distinct levels: the first order intuitive inference and the intuitive inference of higher order—about reasons for the first order intuition. This two levels of inference operate simultaneously. The authors proceed by explaining how the reason module operates under normal circumstances.

Firstly, the authors claim, that the intuitions the reason module provides are not about facts that could be a reason for some unspecified conclusion. Example: "That Amy has a fever is a reason". The reason module provides intuitions which are about facts taken together with the conclusion that they support. Example: "That Amy has a fever is a reason to call the doctor".

Secondly, they claim that in reasoning the output of the reason module is a higher-order conclusion that there are reasons for a lower-order conclusion. Which means that the reason module produces two new conclusions, the second conclusion embedded in the first. The first conclusion is the higher-order argument itself, in other words the metarepresentational intuition that certain reasons support a particular conclusion. Example: "Amy's fever is a good reason to call the doctor now." The second conclusion is the conclusion: Let's call the doctor!, which is embedded in the overall argument and supported by it. The authors call this entire concept the intuitive argument. The intuitive argument is a main function of the reason module. The secondary function of a reason module is the reflective conclusion. The reflective conclusion is a conclusion accepted because of higher-order thinking or reflecting about it. The authors claim that most of our inferences are more intuitive than reflective. Even after they explain how humans reason under normal circumstances there seems that something is missing from their account. The thing that is missing is logic. Mercier and Sperber are very strict here. In reasoning logic is nothing more than a heuristic tool. The authors argue, using Sherlock Holmes as their puppet, that if an argument is valid we generally have no reason to accept it. If an argument is valid and sound, we still generally have no reason to accept it. Only when we have independent reasons which are relevant for us should we accept the argument presented to us. Mercier and Sperber conclude that reasoning is not the use of logic (or of any similar formal system) to derive conclusions. Now the question arises: for what purpose do we use our reason? The authors reject the standard view that we use reason to attain knowledge and make better decisions. They call this view the intellectualist approach. Because this approach fails to explain how humans reason Mercier and Sperber take a different approach, which they call the interactionist approach. So, we use our reason when we are trying to convince others or when others are trying to convince us. The production of arguments proceeds by means of backward inference, from a favored conclusion to reasons that would support it. This function together with mindreading enables us to trust each other and cooperate in a way no other living creature is capable of. But aren't we at risk of being deceived by others? Mercier and Sperber claim that the concept of epistemic vigilance is what keeps us from falling into that trap.

In chapter four, *What reason can and cannot*, Mercier and Sperber continue with their argumentation. They begin by tackling the notion of confirmation bias. They concede that the conformation bias exists, as almost all empirical research confirms this. But they claim that this is nothing to be worried about. The conformation bias is not a bug or a stain in our reasoning, it is actually a feature of our reason. How does the conformation bias as a feature of reason work? Well, the primary function of reason is to convince others and evaluate them when they are trying to convince us. In that sense it is very reasonable to have a conformation bias. Because we enter into this kinds of argumentative processes on a daily basis the authors rename the conformation bias as the myside bias. Mercier and Sperber continue by stating that there are *two faces of reason*. One face is the production of reasons and the other face is the evaluation of other's reasons. This relation is, of course, asymmetrical. In everyday communication we demand honest

answers from others, but we rarely give honest answers ourselves. Most of the time we, more or less subtly, shape our answers to appear better in other people's eyes. The same asymmetry is employed when we produce and evaluate other's reasons. When we produce reasons we are mostly biased for our side and the quality control for our reasons is very poor. On the other hand, when we evaluate other's reasons we are unbiased, we are willing to accept only those reasons which are strong enough and quality control for other people's reasons is very high. This naturally leads to the final segment of Mercier and Sperber's argumentation—working in groups. Firstly, empirical evidence points into direction that working in groups produces better results than working individually. In two studies (Moshman, Moshman and Geil) students were presented with Wason's selection task, first individually then in groups. Individually the students reached a performance of 15% of correct answers and in groups they reached 50% and 80 % respectively. Secondly, working better in groups seems to make theoretical sense as well. If reason is a specific module designed to produce and evaluate reasons, then it is the case that we need each other to reason properly or at least to reason better.

In their fifth and final chapter, *Reason in the wild*, Mercier and Sperber are answering three general questions which can be seen as a possible criticism of their stance.[6] The first question is: is human reason universal? Here, the authors are preemptively reacting to the possibility of their account of human reasoning being WEIRD. The acronym stands for people from Western, Educated, Industrialized, Rich, Democratic countries. Mercier and Sperber conclude that this is not the case. Argumentation can emerge in quite different ways in different cultures but the core of human reasoning stays more or less the same. Empirical evidence shows that very small children are sensitive to argumentation across cultures. The second question is: how do we reason about moral and political topics? Here, the authors argue that most moral and political advancements in history are the product of argumentation. And the third question is: what about solitary geniuses? Mercier and Sperber concede that solitary geniuses exists, but that they are rare and mostly present in rigid fields, like mathematics and logic, in which you can essentially argue with yourself.

Now we should notice some possible points of issue in Mercier and Sperber's argumentation. Firstly, it could be proposed that there should be a stronger connection between human inference and the objective state of affairs in the world. On Mercier and Sperber's account we arrive at the truth almost by accident when we reason with each other. Secondly, the deflation of logic in answering the question how do we reason leaves a huge explanatory gap. In other words, if logic is not the structure of how we reason, then what exactly is logic? The authors are not necessarily obliged to answer this question but this question is one of the outcomes of their account of reasoning. Lastly, argumentation in Mercier and Sperber view is itself deflated to everyday communication between people which in turn presents a serious challenge to the normative aspect of reasoning.

The Enigma of Reason is overall a well written and structurally a precise book. The argumentation is methodically carried out through all of the

[6] These questions are also subchapters of the final chapter *Reason in the wild*.

chapters and the authors are arguing their case extremely well and because of that *The Enigma of Reason* is a book worth reading whether you agree or disagree with the arguments presented there.[7]

DAVID GRČKI
University of Rijeka, Rijeka, Croatia

[7] The writing of this review was supported by the Croatian Science Foundation under grant number IP-06-2016-2408.

Amy Kind (ed.), The Routledge Handbook of Philosophy of Imagination*, London: Routledge, 2016, 482 pp.*

The book is extremely rich. Given that it contains thirty papers, I shall be able to say something only about a part of it. For instance, I shall here for the reasons of space have to skip the rich and interesting Part I, dedicated to historical treatments of imagination, featuring papers on Aristotle (by Deborah K.W. Modrak), Descartes (by Dennis L. Sepper), Hume (by Fabian Dorsch), Kant (by Samantha Matherne), Husserl (by Julia Jansen) and Sartre (by Robert Hopkins).

Let me pass directly to Part II, dedicated to contemporary discussions of imagination. The first paper, "Imagination and mental imagery" by Dominic Gregory stresses the important role of mental images, but adds that "there are, (…), reasons for thinking that sensory mental imagery is not essential to the imagination, (…) The relationships between sensory mental imagery and the imagination are thus notably complex and—like so many of the philosophical areas to which the imagination is central—ripe for further investigation" (107). Bence Nanay in chapter 9, "Imagination and perception" also focuses on image-based imagination. With apologies, I shall skip interesting chapter on "Imagination and belief" by Neil Sinhababu, the one on "Imagination and memory" by Dorothea Debus, one on "Imagination, dreaming, and hallucination" by Jonathan Jenkins Ichikawa and the final one, by the editor of the volume, "Desire like imagination", and recommend them to reader's attention. I also skip part III Imagination in aesthetics, with interesting papers by Nick Wiltsher and Aaron Meskin, James O. Young, Kathleen Stock and Stacie Friend.

So, we now pass to Part IV, "Imagination in philosophy of mind and cognitive science". Let us start from a paper that raises some epistemologically crucial issues. What is the relationship between imagination and rational thought? Are they connected and if yes, what does connect them? Ruth M.J. Byrne, in her "Imagination and rationality" suggests "that counterfactuals provide an important bridge between reasoning and imagination …" (339). The rationale is the following: "Imagination and reasoning depend on the same sorts of underlying computational processes: reasoning depends on cognitive processes that support the imagination of alternatives, and imagination depends on cognitive processes that are based on the same core processes". (Ibid.) She notes that imagination shows an interesting structure: "Remarkably, most people tend to change the same sorts of things in their representation of reality when they imagine an alternative to it, almost as if there were "fault lines" in their representation of reality, junctures or

joints that most people zoom in on to imagine how things could have taken a different turn" (341).

Now, what are the "core processes" that support our imagination of alternatives? Byrne appeals to the notion of mental model that she has been developing for decades together with Philip Johnson Laird. The crucial claim is that "people make inferences by constructing and combining mental models." (348). A mental model is an iconic or analog device, it looks (at least to me) that this iconic character is in many respect analogous to image-based of pictorial nature of many episodes of imagining.

It seems that Byrne supports a strong claim: the thinker does not reason *about* the model using non-iconic devices (a digital language, or…), but rather *in* the model; however, the phrasing is mine, not Byrne's, and she is not very explicit about the matter. However, her work does point in this direction, and this is the way she and Johnson Laird are standardly understood (e.g. by Keith Stenning and Michiel van Lambalgen in their (2008) book, *Human Reasoning and Cognitive Science,* A Bradford book, MIT, ch.10.6). Let me call this view "the pure model-theoretic view of inferences".

Let me note that there is an obvious, more "hybrid" alternative, which sees iconic model as indispensable starting point for reasoning, but leaves space for computational, for instance deductive or propositional, variants of reasoning. Stenning and van Lambalgen argue that the need for such reasoning is present even in the original iconic case, and criticize Byrne and Johnson Laird for not having noticed it. Authors that Byrne herself quotes, J.S.B.T Evans, and D.E. Over, lean more in the same direction. Evans, for instance, describes his models as "epistemic", represented what thinker takes oneself to believe; this allows assignation of degrees of belief to the model, and then thinker's probabilistic reasoning *about* mental models, rather than *in* them. Thomas Metzinger seems to go in the same direction (e.g. in his *Being no one-The Self-Model Theory of Subjectivity* (2004), A Bradford book, p. 62), and the more recent work by Jakob Hohwy (in his *The Predictive Mind* from 2013, OUP) is dedicated to the analysis of hypothetical very rich Bayesian reasoning allegedly done by thinker's cognitive machinery *over* the initial perception grounded mental model(s). (See also Chris Frith (2007), *Making up the Mind How the Brain Creates our Mental World*, Blackwell, p. 126ff.)

The hybrid alternative is slowly winning.

The authors just mentioned do not talk much about imagination. How would the "hybrid" approach answer Byrne's initial question concerning the bridge between reasoning and imagination? One could bring together two kinds of processes: first, the imaginative reasoning grounded in image-like pictorial episodes, and second, the ordinary reasoning grounded in iconic mental models. The second element can be understood either in the pure (Johnson Laird, R. Byrne) way, or in the hybrid way. The two elements seem analogous, and one might suppose that there are causal connections between some episodes of the first, and some of the second element. Suppose, to borrow an example from Tim Williamson, that I ask myself the following vital question: "Can I jump across the river?" I imagine myself jumping from here, or from there, or from a more distant bend. Imagination caters for the singularity of the configuration: this shape of the riverbank

here, the other shape, and so on. I conclude with a counterfactual: Were I to jump from here, I would break my leg.

On the hybrid view, I need information about probabilities, and, equally on the hybrid view, it comes from my spatial-kinematic-and-dynamical mental models. (I might build three models on the spot, using perceptual material plus memory material from my mountaineering experiences). My cognitive mental computer calculates some constraints, Bayesian or others, valid for (on) such models. The models are then stored in my memory, and they will *constrain my imagination* next time when I find myself in a same or similar situation. Imagination feeds models (helps build them) and models, on their part, constrain the exercise of the imagination.

Several papers in the present part concern the role of imagination in understanding other people and our reactions to them. The central phenomenon in play here is mental simulation.

In her chapter "Simulation theory" Shannon Spaulding offers clarificatory and classificatory information that is extremely useful, given that the term "simulation" is used in many senses in the literature, and there is clear need to distinguish them to avoid very bad confusions (her own stance, critical toward the power of simulation, is put forward in her contribution to another collection, due to A. Kind and Peter Kung, (2016), *Knowledge and imagination*, OUP).

Spaulding starts from Alvin Goldman's proposal according to which a process P simulates process P' if and only if first P duplicates, replicates, or resembles P' in some significant respect and two, in its duplication of P', P fulfills one of its purposes or functions. In the case of mindreading simulation, the purpose or function of is to understand target's mental states. She then introduces the crucial distinction between abstract and concrete simulation, the first including activities like computer simulation and the like, and the second being the psychological simulation that involves "sameness of system and fine grained process" (264). The distinction is very helpful, and could save writers from confusions that mark the scene of present-day investigation of simulation.

Spaulding next distinguishes high-level from low-level simulation. She lists three characteristics of the former. First, it "involves imagination in the conventional sense" (267). Second, it explains our engagement with fiction, where we put ourselves "in the fictional character's position and imagine what we would think, feel, and do in that situation. Third, it explains how one can get knowledge through simulational imagination, so that if could be "co- opted to explain how some thought experiments work" (267). In contrast, in low level simulation, "imagination operates unconsciously and automatically."

Let me mention papers which I did not review here. First, the paper on "The cognitive architecture of imaginative resistance" by Kengo Miyazono and Shen yi Liao, that I shall say a few words about in a moment. Next, "Imagination and the Self" by Dilip Ninan, Greg Currie's "Imagination and Learning", Neil Van Leeuwen's "Imagination and action", Deena Skolnick Weisberg's "Imagination and child development" and finally "Imagination and pretense" by Elizabeth Picciuto and Peter Carruthers, and "Can animals imagine?" by Robert W. Mitchell.

Part V is dedicated to imagination in ethics, moral psychology, and political philosophy**.** Next two papers that I want to mention explore the consequences of mind-reading, especially the moral ones. Karsten R. Stueber in her "Empathy and the imagination" stresses the central importance of empathy for gaining access to other minds, and for linking us to our community. She leaves open the question "whether or not empathy has a more foundational role to play in constituting moral normativity" (377).

In his chapter on "Moral imagination" Mark Johnson provides a positive answer to the question and offers a sketch of an extremely pro-imagination oriented theory, of the kind he developed in his *Morality for humans* (2014) book. "Moral imagination", he writes "is our fundamental capacity to imagine how certain values and commitments are likely to play out in future experience…" (263). The quality of our moral thinking is thus "as much an affair of imagination as it is an appropriation of prior knowledge" (Ibid.). A crucial role in moral understanding is thus assigned to one particular kind of imagination, namely simulational, "empathetic" imagination. Indeed, moral deliberation itself is a process of "cognitive conative affective simulation. Simulation "activates emotional responses to the projected situations", thus permitting us "to give voice to various impulses, interests, and values ..." (364).

Unfortunately, Johnson does not clearly tell us how the emotional response to a situation, imagined or real, is related to its value. Empathy with a suffering person can tell me immediately that her situation is bad, of negative value, but this is just the beginning of the story. Is the 'telling' simply recording of a given axiological state, or is it, or some its more idealized version, constitutive of the value? Next, what about non-empathetic simulation? I imaginatively simulate the behavior of sea captains who are saving refugees in the Mediterranean, and I feel admiration for their action, experience it as being of a very high value. What is the relationship between simulation, admiration and value? In the paper, Johnson shows sympathy towards a reductionist account, that would explain value in terms of our coping with the world, but he does not develop it sufficiently for a curious reader.

As far as other prominent ethical alternatives are concerned, Johnson is very critical of what he calls "Moral Law" theory, with its abstract guiding principles. The reader rightly wonders how he would react to the Rawlsian attempt at reconciliation of various sources of moral wisdom. The attempt famously involves three elements: our spontaneous intuitions, the veil-of-ignorance procedure, which is itself simulation-involving (imagine yourself not knowing your particular characteristics, and imagine yourself living in such-and-such arrangement) but not empathetic, and the process of testing principles with the help of spontaneous intuitions. We can imagine that many spontaneous intuitions ("This is, or would be, a terrible deed!") come from empathetic simulation; principles come from neutral, veil-of-ignorance simulation procedure and the testing brings these components together. How close this alternative comes to Johnson's ideal would be a fine thing to know.

Part VI is dedicated to Imagination in epistemology, philosophy of science, and philosophy of mathematics, encompassing the paper by Greg Cur-

rie "Imagination and learning", Roy Sorensen "Thought experiment and imagination", Peter Kung "Imagination and modal epistemology", Adam Toon, "Imagination in scientific modeling" and Andrew Arana "Imagination in mathematics". Let me conclude by mentioning two claims by Sorensen which I find very congenial. First, he proposes that „we can run simulations that have some reliability because our imagination has been shaped by the environment." And notes that experience can improve the fit (431). He also praises modeling, noting that it "improves on the mind's eye level of resolution" and re-organizes our spatial imagination in a positive way. Modeling also overcomes our spatial imagination's bias in favor of alignment and against tilts (426). He is adamant about the role of evolution in shaping our imaginative capacities and making them fit the world. I agree completely.

In short, it is a very impressive collection on the epistemology of imagination, indispensable for anyone who is interested in the methodology of philosophy. I recommend it warmly to the attention of the readers.

NENAD MIŠČEVIĆ
University of Maribor, Maribor, Slovenia

David Miller, Strangers in Our Midst: The Political Philosophy of Immigration, *Cambridge: Harvard University Press, 2016, 218 pp.*

Ethics of immigration has recently emerged as a topic of great interest among political theorists and philosophers. This is not surprising since global migration has recently received a good deal of media coverage and political attention, due to the great number of refugees and economic migrants attempting to enter western liberal democracies.

One of the central questions in the debate on immigration within political theory and philosophy is whether liberal democratic states have a moral right to restrict immigration so as to serve their political, cultural and economic interests best.

The standard position in mainstream international politics regarding this question is based on a widely shared intuition about state autonomy regarding external affairs. Sovereign states have a right to choose their immigration policy as they see fit, with some constrains primarily regarding refugees as a specifically vulnerable group of migrants. This entitles a sovereign state to choose freely whom to accept into their political membership (on what grounds and selection criteria), how to shape their integration policy, and impose both on selected immigrants.

This position has since been widely debated and placed under moral scrutiny. Political philosophers and thinkers question assumptions under-

lying the standard position with regard to human right issues, global justice, equality of opportunity and other recognized liberal and democratic values. Joseph Carens, Chandran Kukhatas, Philip Cole, Kieran Oberman and many others aim to show that the current international immigration regime is deeply unjust, since it often seriously harms interests of prospective immigrants and sometimes, more troubling, their rights.

David Miller is one of the most prominent liberal thinkers firmly positioned on 'the right to exclude' side of the immigration debate. *Strangers in Our Midst: The Political Philosophy of Immigration* summarizes and further develops Miller's previous work in political philosophy with regard to complex issues of immigration.

This book, although rather concise, covers a wide range of important questions in philosophy of immigration: from human rights, cosmopolitanism and global justice, legitimacy and self-determination of states, cultural identity of citizens and social justice, to the moral difference between various types of immigrants regarding their rights and obligations, status and citizenship acquisition.

The book is divided into nine chapters. The first four give outlines of a broader discussion regarding human rights, cosmopolitanism, compatriotism, and the open borders debate, establishing a philosophical basis for further developing of the author`s position. Following the more general introductory chapters, Miller is dedicated to more specific topics, such as the moral difference between refugees and economic migrants, taking into account the debates regarding their rights and integration once accepted into society. In the second part of the book, chapters seven and eight respectively, Miller offers some policy recommendations stemming from the account he develops. The conclusion provides a brief overview of his position with some commentary on methodology.

An important feature of Miller`s account is his commitment to *realism*. Instead of venturing into ideal normative theory, which barely relies on current empirical evidence, Miller grounds his book on a realistic premise. Immigration regimes are under stress, there is a great number of immigrants seeking admission, and citizens' sentiments towards their inclusion are ambiguous and increasingly full of resentment and prejudice (159). "What is needed is clear policy on immigration that can be set out and defended publicly" (160), Miller stresses. Throughout this book, Miller is trying to show that states have a general right to shape their immigration policies freely, which includes the right to restrict immigration, with constrains defined with reference to human rights and values of social justice.

Miller`s position, described by the author himself as "communitarian" and "social democratic" (161), is based on four main values (153): a weak moral cosmopolitanism, national self-determination, fairness and social integration. Each is connected to the topics that Miller tackles throughout the book.

In his discussion regarding cosmopolitanism, Miller is primarily interested in a moral standpoint that establishes equal moral worth of all human beings. This axiom, as it is commonly stated, means simply "that it is unacceptable to respect and treat people differently simply by virtue of some (morally irrelevant) feature such as their gender or skin color" (22), or relevantly for further discussion, their nationality or cultural affiliations.

Weak moral cosmopolitanism, which Miller endorses, relaxes some counter-intuitive implications of the stronger version while appealing to the same value of equal moral worth of all people. According to strong cosmopolitanism, duties we owe to other human beings are the same "regardless of the relationship in which we stand toward them" (22). Pure impartiality in our dealings with other human beings, regardless of any familial, national, civic or emotional ties—that strong cosmopolitanism requires—boils down in the weaker version to a more robust consideration for human rights, leaving space for partiality in relationships that give meaning to human life, like family, friends and for Miller, especially significant, compatriots and co-nationals.

In the discussion on moral cosmopolitanism, Miller builds a case for *compatriot partiality* (21) in external politics of nation states, which includes relevant immigration policies concerning admission, and afterwards, integration.

What weak cosmopolitanism entails is that states should consider the impact of their policies on people outside their borders, including those who wish to enter, thus granting them equal consideration and respect as fellow human beings (22–24). It "tells us nothing about how much weight should be attached to the interests of different groups of people […] affected by a state`s policy, other than that some reason must be given if they are going to be assigned unequal weights" (24).

In our interaction with immigrants, Miller states, we should take their interest and moral standing seriously, choosing policies that protect their basic human rights. Taking interests of immigrants seriously, for Miller, amounts to carefully considering their requests, such as becoming a member of our political community, and giving good reasons to them in case of refusal.

States are thus justified in the differential treatment of their citizens and immigrants, so long as human rights of all are adequately protected. Differential treatment is justified by virtue of the nature of relationships in which we stand to others (26). Special, *associative obligations* that states have towards their citizens arise due to economic, political and cultural relationships within the political and *national* community, which prospective immigrants are *not* (yet) part of. Members of a political community participate together in the economy, the division of labour, legal and political schemes; they relate to one another as citizens, share a similar set of cultural values and a sense of belonging to the distinct community that exists through common history (26). For these reasons, when minimally decent life is guaranteed to all, we cannot expect liberal democracies to set aside special considerations they have towards their citizens in the attempt to accommodate all claims of prospective immigrants. However, what can be expected of states is to treat prospective immigrants with due respect and consideration when their claims, as serious as membership in a political community, are assessed.

The second value that informs Miller's position regarding immigration is *national self-determination* (154). It assumes, according to Miller, a group, *the self*, with a range of values and goals that members recognize as part of their identity (69). For states to be fully self-determining, their

citizens should have a say in shaping the future of their community, according to goals and values recognized as collective. This includes shaping their immigrant policies—whom to accept into the community, when, and under what conditions—since immigration has a profound effect on the political, economic and cultural shape of the community. Self-determination entails "right of a democratic public to make a wide range of policy choices within the limits set by human rights." (62)

The right to control immigration is important for various reasons and Miller points out several, such as cultural identity, control over policies, population size, social trust and the functioning of democracy. He also addresses some objections to his position.

The value of national self-determination, for Miller, is to be supplemented with the idea of territorial jurisdiction, "that implies the right to control the movement of people in and out of that territory" (58).

Appealing to the value of national and cultural identity, and territorial rights, sets Miller apart from other positions in the immigration debate that are mostly based on the value of self-determination. For example, C. H. Wellman`s position is based on the right to freedom of association as an integral part of self-determination of a community.

The discussion on social justice and social practices that Miller brings up when assessing the rights of immigrants and their integration into society, is guided by the *value of fairness* (155). Fairness requires balance and reciprocity in generating social practices, and adequate distribution of burdens and benefits between participants in society. "Balance has to be struck between the claims that immigrants can rightfully make and the responsibilities they can reasonably be expected to assume." (155)

According to Miller, issues of social justice are set apart from human rights issues (116). For a state to be legitimate, it is obligatory to provide protection of basic human rights to everyone present in the state, regardless of their citizenship status. On the other hand, rights of social justice are wider, in the sense that they provide benefits, resources and opportunities that go beyond providing a "minimally decent life", which human rights are designated to protect. For Miller, an adequate immigration policy should be based on reciprocity between the recipient state and the immigrant. Immigrants should be treated fairly, but should also be expected to contribute to the host society and uphold its laws and norms (127). Fairness is constitutive to the notion of citizenship, since the inclusion of legally present long-term immigrants as full political members of society is a prerequisite for upholding liberal and democratic values. Forming a permanent caste of denizens is not in line with the values of a socially just society. This is in line with other prominent thinkers dealing with immigration (e.g. M. Walzer and J. Carens).

Miller applies the notion of fairness to distinguish various types of immigrants. Long term immigrants and temporary immigrants have different sets of rights, obligations and expectations. Treating them fairly, according to Miller, also means respecting designated policies and schemes set with regard to these differences; refugees and irregular immigrants acquire different sets of obligations and policies due to urgency of human rights protection, or irregularity of their presence in the host society. For that

reason, Miller dedicates a more detailed discussion to these types of immigrants—along with their claims, rights and obligations—in relation to their host states.

Finally, closely connected to national self-determination and fairness is the idea of *integrated society* (156). Integration of immigrants into society is essential to forming social relationships between people from different cultural, religious and ethnic backgrounds based on mutual trust, equality and reciprocity. Three different types of integration are acknowledged, with cultural integration being the most controversial. Social integration encompasses participation in social practices, institutions and associations with special emphasis on the way members of different groups interact with each other (132). Civic integration, on the other hand, amounts to full participation in political life of the community, which includes understanding the goals of the community, together with norms and principles that guide political and social life (133). Fostering social and political integration leads to a conflict-free society guided by values of equal access to social services for all permanently residing in the state (134). Cultural integration is rather contestable. It is not, as civic integration is, a requirement, and it can only be an aspiration (149). Critics usually consider it both oppressive and unnecessary (141). Miller is still inclined to consider it as useful aspiration for society that wants to be conflict-free and socially just, since cultural integration means full understanding and acceptance of the public culture, while at the same time it leaves space for accommodating aspects of migrant cultures. If social, political and cultural integration is an expectation that the host society has, with regard to immigrants that wish to permanently become part of the society, then what reciprocity requires is equal inclusion of immigrants in economic, political and social life, with all the benefits that it presupposes (150).

Miller gives an overview of pro-open borders arguments that are compatible with weak moral cosmopolitanism, i.e. arguments from 1) common ownership of the Earth (H. Grotius, I. Kant, M. Risse), 2) global equality of opportunity (J. Carens) and 3) the human right to immigrate (K. Oberman). The purpose of this part of the book, together with the chapter on closed borders, is to show that open borders position can be rejected and that a case for the state's right to restrict immigration can be built and defended.

Open-borders arguments are critically assessed and ultimately rejected, with special concern given to arguments based on the human right to immigrate. A special consideration is dedicated to this specific pro-open border defence strategy, since, without its rejection, Miller`s argument would lose its footing. If it were shown that there is a genuine right to immigrate, restricting immigration with appeal to social justice, cultural and national identity and self-determination would not be possible (162). Human right to immigrate would trump these considerations.

Miller recognizes the three strategies used for justifying the human right to immigrate, which, in a strict sense, is defined as a negative right not to be prevented from entering and settling in any state (49). These strategies are as follows: 1) direct strategy, 2) instrumental strategy and 3) cantilever strategy. The direct strategy aims to show that the right to immigrate has the same grounds as other human rights. What needs to be

shown is that having such a right enables people to live decent lives (50). When discussing the right to immigrate, it is argued that it allows people to enjoy a wide spectrum of *life options* that are not all available in their home country. For Miller, particular preferences and interests, such as career opportunities, relationships or cultural preferences, satisfied only by moving to another country, cannot justify the general human right to immigrate. If states are legitimate, and they protect human rights of its citizens and offer an adequate range of life options important for a decent human life within their borders, then there is no basis for establishing a general right to immigrate that would enable other life options elsewhere. Refugees are a clear example of people whose life options are severely reduced and their human rights limited, which is exactly what gives strength to their claim to immigrate. It is important to note that Miller defends a minimal conception of human rights. Their function is to protect basic human needs and interests and to ensure a minimally decent life. Miller emphasises "that the purpose of human rights is to identify a threshold that must not be crossed rather than to describe a social ideal" (33). This conception of human rights is less demanding than positions that ground human rights in human autonomy, and also see them as positive rights that facilitate the fulfilment of autonomous human life. Thus our particular interest in options outside of our country, when an adequate range of other options are available, is not powerful enough to ground the right to immigrate as a human right.

Instrumental arguments try to show that the right to immigrate is needed as protection of other rights and freedoms (52). Miller specifically stresses the right to exit as recognized in international human rights documents (i.e. Universal Declaration of Human Rights, 1948), which is, however, not matched by a corresponding general right to immigrate. For Miller, the reasons are obvious: we do not need a general right to immigrate for the right to exit to be meaningful. It suffices that one country is willing to allow us to immigrate, and that this can be fulfilled by, for example, bilateral agreements (52). What is needed is an *adequate* level of protection for other rights and freedoms, and for that, Miller stresses, an unlimited right to immigrate is not needed. Instrumental arguments open other important questions, and Miller tackles the so called 'brain-drain' issue; the cost of out-migration of professionals from poor countries, that results in the lack of basic services, such as health service, are borne by the most vulnerable and poor that are not even able to emigrate (53). He concludes, "taking everyone`s human rights into account, recognizing a human right to immigrate does not provide optimal protection for other human rights if one of its effects is to encourage more brain drain from poor states" (53). However, one should note that brain drain is a complex matter, with multidimensional effects on poor states. Using it as philosophical argumentation in the immigration debate, without taking into account complex data from social sciences, can be problematic, as Kieran Oberman points out shrewdly in his article 'Poverty and Immigration policy'.

The third and final strategy for defending the notion of the human right to immigrate is the so-called cantilever strategy. It begins from the notion of the human right to freedom of movement within each state. The right to freedom of movement (or immigration) beyond borders is considered to

be the logical extension of the right to freedom of movement within state borders (53). Joseph Carens advocates this strategy in his book *The Ethics of Immigration* (2013). The reasons for which liberal democracies treat freedom of movement within each state as a general human right are same in the case of movement between states. Opponents of the cantilever argument either have to bite the bullet and concede that there are no violations of rights if people are prevented to move from one region or state to another, both of which provide an adequate range of options, or explain how border control is compatible with the mentioned right (54). Miller gives reasons for destabilizing the analogy between cross-border movement and movement within a state. First, unrestricted freedom of movement is costly. In the domestic case, this is manageable, but on the international scale costs and stakes can be higher, leaving states without mechanisms to protect their citizens. Secondly, the right to freedom of movement in the domestic case prevents the state from dominating minorities, by, for example, resettlements or segregation. Human rights of such groups outside a particular country are not endangered in this way by restricting immigration.

With the open borders positions refuted, and values of national self-determination supplemented with the idea of territorial jurisdiction and the weak cosmopolitanism premise put in place, Miller concludes that it can be shown that states have a right to restrict immigration. The right to close its borders is not coercive, Miller stresses, but preventative, giving regard to self-determination, population control and a proper functioning of democracy. States, however, have responsibilities towards different types of prospective immigrants in admission, and later, for those accepted, in their integration. As noted earlier, these responsibilities are to be guided by values of weak moral cosmopolitanism, fairness, and an integrated society. A large part of the book is dedicated to untangling sets of responsibilities states owe to different types of immigrants. For refugees, these are more extensive and urgent, while for economic migrants, the relationship is of mutual benefit. One group of potential immigrants is singled out, and Miller calls them *particularity claimants* (77). Their claims are "held against one particular state" (77), in virtue of previous events, such as service to the state, or some past injustice suffered by the hand of a particular state. The difference between these is that refugees and economic migrants can hold claims against any state, while particularity claimants have a special relationship with one state against which they hold a claim for admission, for example as a form of reparation.

These and many other questions regarding immigration are raised in this book. The discussion is based on Miller`s previous work on social justice, nationality, global justice and responsibility. His account is informed by important work from other political theorists and philosophers that deal with immigration. However, a more extensive book format would be needed to provide space for a more detailed discussion, with opponents and philosophers of similar accounts—not to mention extensive research of social scientists concerning various aspects of immigration.

Miller does not give his readers a simple recipe for ideal immigration policy. What this book provides are outlines of immigration policy that should be based on four values mentioned earlier. A precise shape of policy

should be developed by each country, in line with other goals (160), values and interests that their communities cherish. Questions raised in this book are not easy, as Miller aimed to show, and faced with crises such as the one in Europe in 2015, scientists, politicians and theorists will be faced with a daring task of answering them.

TAMARA CRNKO
University of Rijeka, Rijeka, Croatia