Philosophy
of Language
and Mind
Network
Conference

26th August – 28th August 2024

BOOK OF ABSTRACTS

KEYNOTE LECTURES

Proper Names and the Referential-Attributive Distinction
Jessica A. Pepp

Uppsala Universitet | jessica.pepp@filosofi.uu.se

In virtue of what does a use of a proper name on a particular occasion (i.e., a linguistic utterance, inscription, token, or particular) refer to whatever it in fact refers to? Saul Kripke (1980) and Keith Donnellan (1970, 1974) sketched different answers to this question. To put things roughly, Kripke takes proper name tokens to refer to what the name names—for Kripke, the thing to which the name was given at the origin of a chain of communication of which the current speaker is a part. Donnellan, on the other hand, takes proper name tokens to refer to the thing that explains a speaker's use of the name on a given occasion— the thing a speaker "has in mind". In Kripke's view, the difference between him and Donnellan over the reference of proper names is resolved by distinguishing *speaker's reference* from *semantic reference.* Donnellan's view is then said to concern the former while Kripke's concerns the latter. A different resolution, which Kripke dismissed, is that not only definite descriptions, but also proper names, have a semantically significant referential use, as well as an attributive use, extending Donnellan's (1966) distinction. Donnellan's view then concerns the referential use, while Kripke's concerns the attributive use. This talk develops this other resolution.

The Enigmatic Nature of Pain: The Problem of Parts and the Problem of Plenty
Sabrina Coninx

Vrije Universiteit Amsterdam | s.coninx@vu.nl

Kripke (1981) famously posited that pain is identified by its 'immediate phenomenological quality', which is contingently connected to the firing of C-fibers. Over the past few decades, these assumptions have been scrutinized and challenged by various scholars, fueled by advances in linguistics, philosophy of mind, and empirical research. In general, pain has thereby moved more and more into the foreground of philosophical discourse. Still, substantial disagreement persists regarding the exact nature of pain, largely due to its seemingly paradoxical and enigmatic characteristics.

My analysis begins with the assumption that broadly construed pain episodes consist of at least four key components: (a) physiological disturbance, (b) phenomenal experience (sensory and affective), (c) action tendencies, and (d) alterations in a broader mental network. Using this framework, we can explore two prominent strands in the current philosophical debate on pain, structured around the 'problem of parts' and the 'problem of plenty'.

The 'problem of parts' concerns the challenge of determining which components provide the necessary and sufficient conditions for a pain episode. It requires deciding which components are essential to pain itself, beyond being merely paradigmatic of pain episodes. This discussion leads us to the field of pain linguistics, which examines the seemingly paradoxical folk concept of pain. Furthermore, this discussion involves a closer examination of the phenomenal characteristics of pain episodes, dissociative cases where some experiential components are apparently absent, and the question of whether a unique 'pain quality' exists that is common to all pains.

The 'problem of plenty' addresses how the relevant components of a pain episode relate to each other. For example, it questions how closely the experiential component is tied to

the presence of a particular physiological disturbance or the elicitation of a specific action. This debate relates to the motivation of strong intentionalist theories, which focus on what pains are 'about'. Here, I will focus on proponents of indicative and imperative variants while considering the causal covariance and biological function of pain. Additionally, I will reflect on the broader issue of how to naturalize pain in light of its complexity and heterogeneity, including insights from current developments in neuroscience.

Finally, I will argue for the future development of a research program focused on the challenges posed by chronic pain. This focus could foster stronger interdisciplinary connections to clinical research and practice, and advance discussions on topics that have received insufficient attention in pain philosophy so far. Emphasizing chronic pain opens new lines of research concerning, for example, the phenomenology of pain experiences and their dynamic generation and maintenance.

Keynote 3
Bullshitting in Public
Marie Guillot
Université de Nanterre | marieguillot@parisnanterre.fr

Contexts in which bullshit (Frankfurt 1986) is rife in public discourse seem to threaten the very possibility of non-defective public argument. I try to understand why that is so. I start by defending a new, non-assertoric account of bullshit which departs from the Frankfurtian approach. In my view, bullshit undermines a prerequisite of all meaningful argument: a common trust that participants will respect the norms governing the practice of arguing. For this reason, prevalent bullshit can cause a *sui generis*, community-wide form of "illocutionary disablement" (Langton 1993): speakers struggle to make their words count as the speech-acts they intend them to be when engaging in public disputes.

Keynote 4
Belief: from Moderate Anti-Realism to Pragmatic Metaphilosophy
Krzysztof Posłajko
Jagiellonian University | krzysztof.poslajko@uj.edu.pl

Traditional debates about the reality of beliefs, especially those arising from the controversies around eliminative materialism, were centred on the question of the existence of beliefs. This ontological way of framing the issue of reality of beliefs has, in my view, tilted the debate against the anti-realist camp: denying the existence of beliefs was widely thought to be self-refuting at worst and counter-intuitive at best. But even if we admit that beliefs do exist (at least in a deflationary sense), there is still a way of questioning their reality: one might question whether belief properties are natural. In this way, a moderate anti-realist view about beliefs becomes the idea that although beliefs exist, they are not natural properties. The issue of naturalness of beliefs can be usefully translated into three more concrete subquestions: of whether beliefs constitute natural kinds, whether they are causally relevant, and whether their content can be naturalised. In all the three cases we have good reasons, even if of negative nature, to prefer the anti-realist option.

The moderate anti-realist position has important implications. First, it supports the autonomy view in the debate about the relationship between folk psychology and cognitive science. If beliefs are best conceptualised as non-natural properties, then it is reasonable to suppose that descriptions in folk psychological vocabulary do not provide us with the

knowledge of the causal structure of the human mind. Rather they offer us a constituting a certain vision of what it is to be human. Causal explanations of human mind, on the other hand, are the domain of middle-level posits of cognitive science. In effect, there should be no place for competition between cognitive science and folk psychology.

The second consequence of moderate anti-realism is its pragmatic approach to metaphilosophy of beliefs. In philosophy of mind, we often deal with issues which can be labelled as the question of the 'boundaries of belief'. We might ask which systems are capable of being genuine believers; e.g. can AIs or non-human animals have them? More often, the question is whether certain mental states of humans (delusions, religious convictions, and implicit biases) should be understood as being cases of beliefs. Adopting a moderately anti-realist view provides basis for the claim that such questions lack objective answers; there are no boundaries of belief which are determined by the world itself. Rather, we should see the concept of belief as plastic: i.e. amenable to changes. This leads to a pragmatic methodology for tackling these questions. We should not be asking questions of the type: is state X really a case of belief, but rather: should we think of state X as being a case of belief? What are the pragmatic benefits/downsides of, e.g. admitting that corporations have beliefs or that religious credences are cases of belief? Moderate anti-realism about beliefs provides us with tools to see many questions in philosophy of mind in a pragmatic fashion.

TALKS

Phenomenal Intentionality and the Sense of Us: A Defense of a Consciousness-First Approach to Collective Intentionality.

Nathan Biglietti

ENS-PSL, France; Institut Jean Nicod | nathan.biglietti@gmail.com

In the contemporary literature on collective mental states, it is often argued that such states present a distinctive intentional structure, irreducible to individual forms of intentionality. This view has been mostly defended in accounting for the distinctiveness of collective actions (Searle 1990). For most authors, such actions are sustained by a specific type of agentive intention, namely "we-intentions", where the "we" is classically ascribed either to their content (Bratman 1999), mode (Tuomela 2007) or subject (Gilbert 1989). More recently, it has been argued that collective emotions possess a "we-mode" intentionality (Salmela 2012). Similarly, it is often argued that joint attention and mutual recognition should be considered as collectively intentional (Gilbert 2011).

In addition, shared mental states are often described as involving a "sense of us" or as being experienced "as *ours*", hence possessing a distinctive phenomenal property. This is especially salient in phenomenological descriptions of collective emotions. For instance, Thonhauser (2018) describe collective emotions as involving a "sense of togetherness", that Léon et al. (2019) take to be what distinguish them from other forms of affective alignment such as emotional contagion. Again, this type of phenomenological description is not restricted to collective emotions. For instance, Pacherie (2014) defend that collective action imply a feeling of "*we-agency*", and Campbell (2005) describe the transition from solitary to joint attention as involving an "*experiential shift*".

The fact that shared mental states possess both a distinctive intentional structure and a distinctive phenomenology raises a straightforward question: *what is the link between these two types of properties*? Both of them have classically been studied separately, within a division of labor roughly corresponding to the division between phenomenologists and analytic philosophers. And this *separatist* tendency is not surprising giving the long-standing tradition in philosophy of mind of distinguishing qualitative states, considered as inherently non-intentional, from intentional states, generally considered as inherently non-conscious (See for instance Ryle 1949). Nevertheless, since the 90's, this view has been largely challenged in favor of a more *intentionalist* picture of the mind (for an overview see Siewert 2002), according to which consciousness and intentionality are intimately related. Endorsing this *intentionalist* perspective, it seems that the sense of us and collective intentionality must be somehow related. In this paper, I will defend one way of accounting for this relation, that I will call *consciousness-first approach* to collective intentionality.

Even if the question of the relation between collective intentionality and phenomenology haven't been directly addressed, we can identify two antagonistic approaches among authors who tried to account for the sense of us. Some seems to endorse an *Intentionality-first approach*, and try to ground this experiential property in a collective intentional state it is supposed to emerge from. For instance, Salmela (2022), argues that the "sense of togetherness" emerges from the shared concerns underlying collective emotions. On that approach, phenomenal properties play no role in accounting for collective intentional states, but should be rather considered as a mere by-product of those. Others, however, plead for a *Consciousness-first approach*, according to which collectively intentional properties are actually grounded in phenomenal ones. For instance, Schmid (2014, 2023) suggests that collective intentionality arises from a *plural* pre-reflective

self-awareness, or Crone (2018, 2021) argues that a sense of us pre-structure explicit we-attitudes. On this second approach, the sense of us is what grounds collective intentionality, and is considered as necessary for collective mental states. Furthermore, the sense of us plays for these authors a central theoretical function in allowing them to avoid infinite regress in accounting for the collectivity of shared mental states.

While I am more sympathetic to *consciousness-first approaches*, I suggest that it is by no means clear in what sense an experiential property can be said to ground collective intentionality. In this paper, I propose to defend such views by clarifying the relation between consciousness and intentionality, and suggesting that collective mental states obtain their intentional structure in virtue of their phenomenal character. More precisely, I'll argue that the intentional specificities of collective mental states fall into what Horgan & Tienson (2002) call "phenomenal intentionality", in other words intentional properties that are constitutively determined by phenomenology, such that the sense of us, to rephrase Kriegel (2014), is what injects collective intentionality into the world. I will first present phenomenal intentionality theories and the three main arguments in their favor (Woodward 2019). First, *assesability for accuracy* arguments (Siewert 1998) suggest that a subject can be assessed for accuracy in virtue of the phenomenal character of his or her perceptual states. Second, *phenomenal duplicates* arguments (Loar 2003) suggest that we cannot conceive a pair of perceptual states that are phenomenologically identical but differ in all their intentional properties. And third, *intentional contrast* arguments (Siegel 2010) suggest that we cannot conceive a pair of perceptual states that differ intentionally without differing phenomenologically. According to their defender these three facts are best explained by perceptual phenomenal intentionality.

I will then show that all of these three arguments can be applied to the sense of us. In other words, I will defend that one can be assessed for accuracy in virtue of his or her experience of togetherness, that two identical experiences of togetherness that nevertheless differ in the presence or absence of co-participants both count as collectively intentional, and that a transition from individual to shared mental states cannot happen without a phenomenal switch to an experience of togetherness. I will argue that the sense of us should therefore be considered as a phenomenal-intentional property that grounds the intentional distinctiveness shared mental states. Finally, I will discuss two potential objections to this view: one regarding the nonconscious character of plural subjects, and a second concerning the absence of qualitative character in shared propositional attitudes. I will then conclude by stating two consequences of this view, one regarding the necessity of taking the sense of us as a distinguishing criterion for shared mental states, and a second regarding the constrains that phenomenological description of collective mental states should impose on accounts of collective intentionality.

Tuesday, 27/Aug/2024 18:10 – 18:50

Philosophy of Language

Conversational Implicit Bias and Meta-Linguistic Stereotypes

Stina Björkholm

The Institute for Futures Studies, Stockholm, Sweden | stina.bjorkholm@iffs.se

People who explicitly endorse progressive and egalitarian values sometimes act in prejudiced ways towards members of stigmatized groups. This is often explained by appeal to implicit biases. The thought is, roughly, that people may *explicitly* endorse anti-racist and feminist beliefs, and yet have *implicit* attitudes that contradict these beliefs. This

psychological understanding of implicit bias has been criticized by those who maintain that it neglects the *structural* aspects of the phenomenon. However, a problem with structural explanations is that they are more difficult to grasp; it is unclear what the nature of structures are and whether their existence goes beyond individuals – and if so, how?

In this paper, I offer a structural pragmatic account of the mechanisms behind what I call *conversational* implicit bias. These include the biases that members of stigmatized groups themselves experience in conversational interactions and the ways that utterances about members of stigmatized groups enforce biases. For instance, consider the case of a mother who returns from parental leave and the first thing her colleague says is "Are you back *already*?". We can imagine that this colleague explicitly endorses feminist values, but nevertheless makes this type of utterance which underscores a stereotype that women should prioritize family over career. Another example is Fricker's influential work on testimonial injustice, where the central idea is that a speakers' utterance is assigned less credibility because of a prejudiced stereotype about her social identity held by the listener (Fricker, 2007, p. 2). The account of implicit bias in conversation presented here will challenge the last part of this characterization of testimonial injustice, calling into question what it means for the stereotype to be *held by the listener*.

The account of conversational implicit bias presented builds on dynamic pragmatics, which is broadly the study of the assumptions that interlocutors make about their communicative exchange, such as the meanings of the words they use and the purposes of their conversation. One of the most influential versions of this is Stalnaker's view that interlocutors make *pragmatic presuppositions* about what is mutually believed. The set of propositions that are mutually accepted to be common belief form the *common ground*. This way of modelling a context can be understood as a rational reconstruction of what happens in conversations. When a proposition is included in the common ground, this means that the proposition is henceforth *accepted* as common belief, where 'acceptance' is understood as a sociolinguistic disposition to behave in conversation as if the proposition is mutually believed to be true. Interlocutors can accept a proposition *for the sake of conversation* – without actually believing it – in which case they will be expected to adjust their behavior in accordance with it (Stalnaker, 1999, pp. 52, 58).

By drawing from Levinson's (2000, 37) pragmatic view, I argue that interlocutors accept *meta-linguistic stereotypes* as default presuppositions in the common ground. For instance, because interlocutors mutually accept the stereotype that *tomatoes are red*, they tend to say things such as "We need tomatoes" not "We need *red* tomatoes" because it is mutually accepted that the object described instantiates the stereotype, unless one suggests otherwise.

The main hypothesis pursued in this paper is that since stereotypes consist of generic propositions, they inherit two problematic features of generic propositions that are about members of social groups.

First, according to Leslie (2015), we tend to think of social kinds as *essentialized* which means that we attribute to them a shared underlying nature. Haslanger (2012) argues that generics such as "Women are submissive" or "Boys don't cry" implicate that there is something about the nature of women that make them submissive and something about the nature of boys that they do not cry. Meta-linguistic stereotypes about social identity groups inherit this tendency to essentialize from the generic propositions that they consist of.

Second, generics about social groups have a *normative* force (Leslie, 2015). Roughly, the thought is that a generic such as "Boys don't cry" says something about how boys *ought* to behave, not only how they *do* behave. According to Leslie, the normative force associated

with generics such as "Boys don't cry" or "Women are submissive" stem from the *social roles* that we associate with some social kinds (such as boys and women) and that members of this social kind have a *prima facie* obligation to possess the features that are constitutive of the social role (Leslie, 2015, p. 130).

This normative, imperative-like, force of generics about social identity groups affects the nature of meta-linguistic stereotypes about them. This can be captured by expanding the notion of context to include not only propositions in the common ground, but also actions on *to-do lists* (Portner, 2004, 2007). Very roughly, the to-do list function assigns properties onto each interlocutor's to-do list which represents the actions that they are expected to perform. The normative dimension of social identity stereotypes can be understood as the acts that are considered appropriate or even obligatory for a member of that group to perform. As such, in a context where the interlocutors accept the social identity stereotype that *boys do not cry*, the property of *not crying* will be assigned by default to the to-do list of boys that enter that conversational context. Since social identity stereotypes include items on to-do lists, they affect how interlocutors expect others to behave based on their social identities. If an individual acts in a way that diverges from these expectations, s/he will be considered to act inappropriately. For instance, if the context includes the stereotype that women are submissive, then *being submissive* will be assigned by default onto female interlocutors' to-do lists.

Meta-linguistic stereotypes provide an important tool in giving a structural account of implicit bias since they constitute shared sociolinguistic dispositions rather than individual mental states. Hence, testimonial injustice does not only arise when the listener *believes* a prejudiced stereotype, but it also when the interlocutors *accept* such stereotypes in the discourse context – even if neither of them explicitly believes it.

Meaning by Unintentional Permission
Constant Bonard
University of Bern, Switzerland | constant.bonard@gmail.com

Sam and Maria are walking in a park when Maria slips on a banana peel and falls. Sam laughs. A silence ensues as Maria gets up. Maria is angry. Sam had no intention of conveying any information to Maria by behaving in this way. Nevertheless, by his laughter and lack of apology, Sam implies certain things. In particular, this behavior, combined with the knowledge that Maria and Sam already had in common, allows for the following interpretation:
(p) Sam lacks respect for Maria.[1]

If we ask Sam: "What did you mean by that laugh? And by that silence?", he could legitimately answer: "Nothing! I didn't mean anything. I didn't want to communicate anything." We should think of this example in such a way that this answer is neither false nor misleading.

Nevertheless, even if he did not mean to say anything, by his behavior, Sam somehow *implies* that (p). In other words, by laughing and failing to apologize, Sam involuntarily sends the message that (p). Maria can legitimately reproach him for this and can, moreover, expect Sam to know that she can reproach him for this even before she tells him so. I will say that, with this behavior, Sam *means (p) by unintentional permission* or, equivalently, that he *unintentionally allows his behavior to mean that (p)*. Indeed, although it is unintentional,

with his behavior Sam allows, permits, or legitimizes certain interpretations by Maria, including (p). On the one hand, he had no intention of conveying this message but, on the other hand, he manifestly could have prevented this interpretation. Because he could, he gives Maria reason to interpret him as conveying certain messages, including (p).

In this article, I present an analysis of this phenomenon of meaning by unintentional permission. I will do so by presenting a model inspired by the work of Grice and his heirs that I call "the extended Gricean model", abbreviated EGM. More specifically, I will defend the following three points:

(1) Theories of implicatures developed by Grice (1989) and his heirs (for a review, see Davis 2024) fall short of being able to analyze meaning by unintentional permission insofar as these theories focus on intentional meaning (or ostensive-inferential communication).

(2) Some adaptations of the Gricean framework allow us to account for these cases. This leads me to present the EGM and define its key notions, which are adapted from Gricean notions: *meaning by unintentional permission* (which replaces "utterer's meaning" in the EGM), *to allow x to F* (which replaces "to intend x to F" in the EGM), and the *principle of rationalization of permission* (which replaces the Gricean Cooperative Principle and, more generally, the principle of rationalization of Kasher (1982)).

(3) Finally, I argue that the EGM can explain many interpretations that we make on a daily basis that escape the existing (post-)Gricean theories as well as the other communication theories that I know of. We will see in particular how this model explains why, for example, Sam is responsible for what he implies by his behavior, even if he does not mean anything by his behavior. This is because, in a nutshell, he has some control over what he allows his behavior to mean.

To give a foretaste of the EGM, here are the definitions of the key notions mentioned above:

Meaning by unintentional permission

By the production of the stimuli x, the sender S means that p by unintentional permission to the receivers R if, and only if,   S unintentionally allows x

(i) to generate the mental effects e in R – typically the relevant effect is *to make p manifest to R* – and

(ii) to make (i) mutually manifest to S and R.

Note: Specifying e is tantamount to specifying p, i.e. the content of meaning by involuntary permission.

To allow x to F

The sender S allows the stimuli x produced by S between the instants t0 and t1 to F if and only if

- S has control over whether S Fs between t0 and t1 and

- it is manifest to S between t0 and t1 that S may F with x.

Note: "S has control over whether S Fs" can be analyzed in at least two ways: Either we say that it is *within S's power* not to F (Chisholm 1967) or F-ing is produced by a mental mechanism of S that is *sensitive to reasons* (cf. the notion of 'moderate guidance-control' (Fischer and Ravizza, 1998, ch. 3)). In any case, S can be taken to be responsible for F-ing.

Principle of rationalization of permission

If S allows in a mutually manifest way x to generate e in R, then one may assume that these effects are evaluated by S as being conducive to S's goals, *ceteris paribus*. In particular, one may assume that S would not have allowed x to generate e in R if this had been evaluated as less conducive to S's goals than not allowing it, *ceteris paribus*.

Note: "to evaluate" is here to be understood as a basic, rapid, non-demanding cognitive process (see Moors, 2022, ch. 6-7).

References

Chisholm, R. M. (1967). He could have done otherwise. *The Journal of Philosophy*, *64*(13), 409–417.

Davis, W. (2024). "Implicature", *The Stanford Encyclopedia of Philosophy*, Zalta, E. & Nodelman, U. (eds.), <https://plato.stanford.edu/archives/spr2024/entries/implicature/>.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Kasher, A. (1982). Gricean Inference Revisited. *Philosophica*, *29*(1), 25–44.

Moors, A. (2022). *Demystifying emotions.* Cambridge University Press.

[1] Or something like that. The precise content of (p) is not relevant to my argument. It should be understood in a flexible enough way to be replaced indifferently by: Sam does not respect Maria at her true value; Sam does not treat her in a sufficiently respectful way; Sam respects her less than she would like; etc.

Wednesday, 28/Aug/2024 11:30 – 12:10

Philosophy of Mind

Defining Cognitive Permeation

Peter Brössel

Ruhr University Bochum, Germany | peter.broessel@rub.de

This paper seeks to provide a rigorous definition of the concept of Cognitive Penetration (Pylyshyn 1999) or, as it is nowadays, more often called Cognitive Permeation (CP) (Burnston 2017). Alleged cases of CP include the perception of brightness in faces based on beliefs about their race and skin colour (Macpherson 2012), the perception of colour based on beliefs about the typical colour of objects (Witzel 2016), and the perception of facial features, based on beliefs about their emotions (Marchi & Newen 2015). The debate whether perception is permeated by cognition is ongoing (Toribio 2015, Firestone and Scholl 2015, Newen & Vetter 2017). One of the reasons is that we do not have a sufficiently precise definition of CP

In the literature, there are two approaches to understanding CP. On the one hand, the neurobiological approach conceptualises CP as a (sufficiently direct) causal influence on a specific segment of visual processing known as Early Vision. According to this approach, whether perception is permeated by cognition depends on whether there are direct neuronal links from cognitive segments to early vision segments. Approaches from philosophy and perceptual psychology, on the other hand, focus primarily on the kinds of mental representations involved and on whether the influence from cognitive states on perceptual states, can be ascribed to the content of these states or other contributors such as attention, etc. This paper argues that the latter approach to defining CP is superior. Only this approach to defining CP is relevant for theories of perceptual justification (Siegel 2012).

Based on the selected approach then, the study commences with defending four criteria of adequacy for a definition of CP.

1. The causal connection between the cognitive and perceptual mental representations is internal and mental.

2. There must be a semantic connection between the cognitive mental representation and the perceptual mental representation.

3. The causal connection is due to a computation over the semantic content of the cognitive and perceptual mental representations.

4. CP is not Theory-Ladenness of Perception.

These criteria are motivated by discussing examples from the literature (Macpherson 2017, Stokes 2015) of causal influence from cognition to perception. Each of these examples also shows that previous proposal for defining CP are inadequate because they do not satisfy all of the above criteria of adequacy for a definition of CP.

This discussion of the criteria of adequacy ends with a dilemma that explains why previous definitions of CP were inadequate. If the content of perceptual mental representation were identical to the conceptual content of belief-like cognitive mental representations, then CP and the theory-ladenness of perception could not be distinguished sufficiently. Suppose the contents of perceptual mental representation were sufficiently different from the conceptual contents of beliefs, i.e., by being non-conceptual, as many philosophers of perception suggest. In that case, previous approaches to CP and the philosophy of perception generally cannot explain how a semantic relationship between belief-like cognitive mental representations and non-belief-like perceptual mental representations can exist.

To resolve this dilemma and to offer an adequate definition of CP, the paper provides an account of the non-conceptual content of perceptual mental representations and its semantic relations to the conceptual content of beliefs. The account is a generalisation of my previous account of the semantic relationships between perception and beliefs (omitted) which is also applied to understanding the interface between the non-conceptual content of sensorimotor representations and the conceptual content of intentions (omitted). The core idea is that the non-conceptual content of perceptual mental representations can be captured in terms of perceptual similarity spaces. Perceptual concepts, i.e., the building blocks of belief-like mental representations, can be analysed semantically as regions in such a similarity space. For example, the semantics of the colour concept RED can be understood in terms of a specific region in the colour space. The action concept KICK can semantically be analysed as a region in the sensorimotor space representing bodily positions and movements (see Gärdenfors & Warglien 2012). This approach puts us in the position to satisfy the second and third criterion of adequacy of definitions of CP without conflating CP with the theory-leadenness of perception and thus violate criteria 4. Based on this account, the paper suggests the following definition of CP (here, it is only outlined).

A perceptual mental representation P R is cognitively permeated if and only if there is a cognitive mental representation C such that:

1. There is a semantic connection between the content CNPR of PR and the content CNC of C within the relevant perceptual similarity space along the lines characterised above.

2. In virtue of the causal and semantical connections between PR and C, respectively CNPR and CNC, variations in the conceptual content of C would lead to similar variations in the non-conceptual content of PR.

Subsequently, the paper shows that this definition satisfies the four criteria of adequacy for definitions of cognitive penetration.

Literature

Burnston, D. (2017). Cognitive penetration and the cognition-perception interface. Synthese 194, 3645–3668.

Firestone, C., and B. Scholl. (2015). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. Behavioral and Brain Sciences.

Gärdenfors, P., Warglien, M. (2012). Using Conceptual Spaces to Model Action and Events. Journal of Semantics 29, 487–519.

Macpherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. Philosophy and Phenomenological Research, 84(1), 24–62.

Macpherson, F. (2017). The relationship between cognitive penetration and predictive coding. Consciousness and Cognition 47, 6–16.

Marchi F, Newen A. (2015). Cognitive penetrability and emotion recognition in human facial expressions. Frontiers of Psychology 19, doi: 10.3389/fpsyg.2015.00828. PMID: 26150796; PMCID: PMC4473593.

Newen A, Vetter P. (2017) Why cognitive penetration of our perceptual experience is still the most plausible account. Consciousness and Cognition 47, 26–37.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. The Behavioral and Brain Sciences, 22(3), 341–365; discussion 366–423.

Siegel, S. (2012). Cognitive penetrability and perceptual justification. Nouˆs, 46(2), 201–222.

Stokes, D. (2015). Towards a consequentialist understanding of cognitive penetration, eds. J. Zeimbekis, and A. Raftopoulos 2015, 75–100.

Toribio, J. (2015). Visual experience: Rich but impenetrable. Synthese. doi:10.1007/s11229-015-0889-8.

Witzel, C. (2016). An Easy Way to Show Memory Color Effects. I-Perception, 7(5). https://doi.org/10.1177/2041669516663751

Quine's Response to Davidson's Cartesian Reading of Quine's "Proximal' Account of Meaning

Agnese Casellato
University College Dublin, Ireland | maria.casellato@ucdconnect.ie

Donald Davidson has argued that W.V. Quine's empiricism depends on sensory stimulations as epistemic intermediaries between theory and world (1974, 1990), a view which should be rejected. Once this step is taken, the doctrine that empirical beliefs play a foundational epistemic role should also be abandoned. Nothing distinctive then, Davidson has suggested, is left to call empiricism. This paper addresses what Davidson describes as the 'Cartesian' spirit of Quine's account of observational beliefs (1990). I shall argue that Davidson's reading may not reflect what Quine's approach amounts to and that it may, perhaps, even conflict with it.

Davidson suggests that Quine's empiricism involves a Cartesian epistemology based on sensory intermediaries, much like traditional forms of empiricism are based on sense data (2001). In this paper, I shall first focus on analysing what Davidson means by qualifying Quine's empiricism as Cartesian. This is important as Quine is rarely associated to first-personal approaches to knowledge or meaning. Given the anti-mentalistic strictures imposed by naturalism, such as the requirement to explicate the empiricist notion of objects 'before the mind' (1989) in terms of sensory exteroceptors, neural firings or nerve endings, Quine's epistemology is customarily associated with third-personal scientific accounts of language. In a similar vein, Davidson does not think that Quine reduces the subject matter of observation sentences to patterns of sensory stimulation. Davidson's reading is significantly more nuanced.

In Davidson's view, though shying away from traditional empiricism, Quine's use of the notion of stimulus meaning in *Word and Object* (1960) leads to similar consequences. The core of Quine's empiricism consists in what Davidson calls the "proximal" theory of meaning and belief, which establishes that two sentences mean the same thing if they are prompted by the same patterns of stimulation (Davidson, 1990: 71-73). According to Davidson, stimulus meaning introduces a gap between the radical translator's conception of the speaker's beliefs on one hand, and the world with its public objects and events on the other

hand. While sense data philosophers would have placed such gap within the subject's mind, Quine instead has externalized it to an aspect of the world potentially describable in scientific terms. Nevertheless, Quine's naturalization of sense data leads to the same subjectivist consequences: it is by asking herself which patterns of nerve endings would prompt a speaker to accept or reject an observation sentence, and then by translating that sentence into one that she would be disposed to accept or reject, that the linguist is able to construct a translation manual (1994). Since different patterns might be stimulated in different people by the same situations, the same utterance might be assigned conflicting truth values in the mouths of different people. In Davidson's eyes, precisely because the linguist's or speaker's beliefs are independent from sensory intermediaries, and both are independent from the public world, an utterance of theirs of the same expression would issue in conflicting truth values even assuming that their response was caused by exactly the same token-stimulations or proximal conditions of assent and dissent.

This paper then moves on to an assessment from Quine's perspective of Davidson's Cartesian interpretation. The conclusion I attempt to support is twofold. On one hand, though aware of the perils of assimilating Quine's empiricism to a mentalist account of sensory stimulation, I argue that the subjectivist aspect which Davidson attributes to mentalist kinds of empiricism (1989) may constitute one possible way to make sense of Davidson's view of the consequences of Quine's empiricism, namely that it entails skepticism and a certain kind of relativism (1990). What I mean by subjectivism is the view that it is the speaker's circumstantial and idiosyncratic theory of how the world appears to her which governs the content and meaning of her empirical beliefs. On the other hand, I attempt to argue that Quine's empiricism does not fit the design of subjectivism. To do so, I examine it in light of Quine's thesis of the reciprocal containment of ontology and epistemology (1981: 72).

A consequence of the containment thesis is that all existential claims obey the same standards of evidence, namely those conducive to more science. From the evidential standpoint, therefore, all objects are on equal footing: it is only arbitrarily that sensory stimulations are more fundamental than other objects. By following through Quine's naturalism, empiricism emerges as, at once a scientific description of an aspect of the world such as language learning and theory construction, and a validation of such description from within the purview of our scientific worldview. Any air of circularity may be tolerated on the grounds that there simply is no "first philosophy" (Quine, 1981: 77), no domain of assessment reaching beyond science and against which science can be compared. The containment thesis I argue, supports two conclusions. Firstly, it at least casts doubt on the viability of attributing to Quine the assignment of a privileged role to sensory intermediaries. Secondly, it has significant bearing on Quine's use of the notion of stimulus meaning, which notion therefore is of no immediate use for the theory of meaning. Empiricism belongs to a different level of Quine's philosophical system, namely Quine's naturalized epistemology. Instead, because of his genuinely semantical concerns, Davidson reads it as a theory of translation; but instead, it should be read as a second-order scientific claim without immediate bearing for the semantic work of the linguist.

Davidson's critical involvement with Quine on empiricism has significantly informed Quine's philosophical development from the 1970's to the 1990's. Davidson's arguments sometimes prompt Quine to reconsider certain aspects of his earlier views. For instance, Quine has rejected the thesis of a "homology between people's nerve endings" (Quine, Davidson, 1994: 227). Other times Davidson prompts Quine to provide new explanations for older commitments. This is where empiricism should be located. Quine is unrepentant in retaining the philosophical role of stimulation patterns (Quine, 1981: 38). However, rather

than providing a semantic criterion for translation, stimulus meanings provide a definition for why it works (Quine, 1999: 74).

References

Davidson, D., (1990), 'Meaning, Truth and Evidence', in (1990) *Perspectives on Quine*, Barrett, R., B., Gibson, R., F., (eds.), Blackwell.

- (2001a) 'Comments on Karlovy Vary Papers', in (2001), *Interpreting Davidson,* Kotatko, P., Pagin, P., Segal, G., (eds.), Stanford University.

Quine, W., V., (1960b) *Word and Object*, MIT Press, 2013.

- (1981) *Theories and Things*, The Belknap Press of Harvard University Press.

- (1999), 'Where do We Disagree', in *The Philosophy of Donald Davidson*, Hahn L. E., (ed.).

Quine, W., V., Davidson, (1994), D., 'Exchange between Donald Davidson and WV Quine following Davidson's lecture', *Theoria*.

Monday, 26/Aug/2024 12:10 – 12:50
Philosophy of Language
Content-Bearers and Indexicality

Tadeusz Ciecierski
University of Warsaw, Poland | taci@uw.edu.pl

One potential conservative reaction to the concept of indexicality of content is a theory that treats indexicality as a property inherent to content-bearers. While this view aligns intuitively with linguistic cases, as indexicality is commonly seen as a property of expressions, it requires further elucidation when applied to intentional (mental) states as content-bearers. In this paper, I aim to present a theory that elucidates the indexicality of attitudes and other mental states by examining the properties of the bearers or vehicles of content.

The theory I shall present departs from an account of indexicality rooted in Frege's philosophy, specifically a *hybrid expression view* (cf. Frege, 1956). Various versions of this view (cf. Künne, 1992; Künne, 2010; Textor, 2007, 2015; Kripke, 2008; Penco, 2013; Ciecierski, 2019) share the common feature of conceiving content-bearers as complex objects. These objects, in addition to narrowly conceived components (linguistic expressions in the case of utterances or mental representations/forms in the case of mental states), consist of contextual elements such as the speaker, time of utterance, or place of utterance. Recognizing the differences in these content-bearers allows for the expression of a single standard and non-indexical content through appropriately contextually coordinated but different content-bearers, while different non-indexical contents are expressible by uncoordinated content-bearers.

However, the theory is not immune to objections (cf. Perry, 1977, 491; Kaplan, 1989, 538). A notable challenge arises in explaining the intuition that an individual who loses track of time, like Rip van Winkle, shares the same thought when considering that today is sunny (on a day d in 1789) and when thinking that yesterday was sunny (on a day d' in 1800). One response to this challenge (cf. Tichy, 1986, 40; Textor, 2011, 168) suggests that a person who loses track of time fails to accurately capture the content of the thought that yesterday was sunny (on a day d' in 1800). This response, however, rejects the initial intuition entirely and flies in the face of the idea of the transparency of mental content: the content of Rip van Winkle thought is different from the content he believes his thought has.

To address this problem, I propose the internalistic version of the hybrid expression view of indexicality. While the standard hybrid expression view adopts an externalistic perspective, claiming that a hybrid expression (content-bearer) consists of a vehicle and an aspect of the relevant externally existing context of utterance (*Externalistic Hybridity Claim*), the internalistic version posits that a hybrid expression consists of a vehicle and an aspect of the relevant *mental representation* of a possible context of utterance (*Internalistic Hybridity Claim*). Accepting Internalistic Hybridity enables a nuanced treatment of cases where individuals are "lost in time." For instance, Rip van Winkle, while entertaining the thought that yesterday was sunny, grasps the content of the hybrid representation consisting of a vehicle (the mental counterpart of "yesterday") and the representation of the time of utterance corresponding to d+1 1789. Consequently, he fails to entertain the content that it was sunny on d'-1 1800, but he apprehends a different yet related content. His mistake in content attribution is a result of a factual error—selecting the wrong context as actual. This analysis aligns with the intuitive transparency of mental content: Rip van Winkle grasps the appropriate content and maintains a disposition to differentiate between it and the content expressed had the time of utterance been located on a specific day in 1800.

In the final parts of my paper I shall discuss the challenges that the theory might face as well as its potential merits. One potential merit of this theory is its uniform treatment of linguistic and mental indexicality. One potential challenge is the tension between the *Internalistic Hybridity Claim* and the directly referential nature of indexicals. I shall argue, however, that there is no conflict between the two: *Internalistic Hybridity* does not entail that indexicals and demonstratives are not directly referential, it rather claims that in cases where the mental representations of relevant aspects of the context do not correspond to actual aspects of the context the reference of indexicals and demonstratives might be moved to a different object (while remaining direct in the Kaplanian sense).

References

Ciecierski, T. (2019). Hybrid Expressions. In: Bella, G., Bouquet, P. (eds) *Modeling and Using Context*. CONTEXT 2019. Lecture Notes in Computer Science, vol 11939. Springer, Cham.

Frege, G. (1956). The thought: A logical inquiry. *Mind* 65 (259):289-311.

Kripke, S. (2008). Frege's Theory of Sense and Reference: Some Exegetical Notes. *Theoria* 74 (3):181-218.

Kaplan, David (1989). Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes From Kaplan*. Oxford University Press. pp. 481-563.

Künne, W.(1992). Hybrid proper names. *Mind* 101, 721–731

Künne, W. (2010). Sense, reference and hybridity. *Dialectica* 64, 529–551

Penco, C. (2013). Indexicals as demonstratives: on the debate between Kripke and Künne. *Grazer Philophische Studien* 88, 55–73

Perry, J. (1977). Frege on Demonstratives. *The Philosophical Review*, *86*(4), 474–497. https://doi.org/10.2307/2184564

Textor, M. (2007). Frege's theory of hybrid proper names developed and defended. *Mind* 116, 947–981

Textor, M. (2011). *Frege on Sense and Reference*, Routledge

Textor, M. (2015). Frege's theory of hybrid proper names extended. *Mind* 124, 823–847

Tichy, P. (1986). Frege and the case of the missing sense. *Grazer Philophische Studien* 27(1), 27–47

Language Logicality and the Logic of Identification

Giada Coleschi

University of Genoa, Italy | giada.coleschi@gmail.com

According to the logicality of language hypothesis (hereinafter, LLH), logical considerations are relevant for syntactic formation to the point that they are needed to explain some ungrammaticalities (Gajewski 2002, 2009; Fox & Hackl 2006; Chierchia 2013, 2021; Abrusán 2014; Del Pinal 2019, 2021). In other words, the syntactic unacceptability of some linguistic constructions is traced back to their logical status, suggesting that speakers judge a word-sequence as ungrammatical or ill-formed when it is always false or always true. The aim of this work is to examine the recent literature that attempts to offer alternative explanations and test the solidity of the objections it proposes. In particular, we will look at cases concerning identity and co-binding.

Chief evidence in support of LLH comes from word-sequences, such as (1)–(4), judged ungrammatical *qua* contradictory.

(1) *Some students but John passed the exam

(2) *There are any cookies left

(3) *How fast didn't you drive?

(4) *How did John regret that he behaved at the party?

Other word-sequences, such as (5)–(7), are judged ungrammatical *qua* tautological.

(5) *There is every fly in my soup

(6) *Mary is taller than no student is

(7) *At least zero students smoked

However, in general, contradictions and tautologies are not ungrammatical.

(8) This table is red and it is not red

(9) It is raining or it is not raining

Within LLH, the syntactic acceptability of (8)–(9) is not questioned. As a result, the asymmetry between the ungrammaticality of (1)–(7), due to their logical status, and the grammaticality of (8)–(9), notwithstanding their logical status, has to be accounted for. This is the so-called «analyticity puzzle».

According to an early articulation of LLH (Gajewski 2002, 2009; Fox & Hackl 2006; Chierchia 2013), one must assume a purely linguistic logic and associate linguistic constructions with particularly austere representations, which reveal truly contradictory or tautological contents only in the case of constructions later judged ungrammatical. According to an alternative articulation, on the other hand, contradictions and tautologies are ungrammatical insofar as it is not possible to make them informative, by modulating their lexical content (Del Pinal 2019, 2021; Pistoia-Reda & Sauerland 2021; Pistoia-Reda & San Mauro 2021).

There have been attempts to explain the rejection of (1)–(7) typologically, i.e., in terms of a failure of composition (Abrusán, Asher & Van de Cruys 2021). To that extent, (1)–(7) would be similar to examples of semantic anomaly, such as (10)–(11).

(10) #Tigers are Zermelo-Fraenkel sets

(11) #My toothbrush is pregnant

According to LLH, (1)–(7) are only superficially uninterpretable: when analysed, they receive an interpretation, which is that of being either contradictory or tautological. According to Abrusán, Asher & Van de Cruys (2021), (1)–(7) are instead uninterpretable for, in building up their semantic representation, an insuperable semantic problem is encountered.

The critical objective of Abrusán, Asher & Van de Cruys (2021) is primarily the early articulation proposed by Gajewski (2002, 2009). However, according to them, the alternative articulation proposed by Del Pinal (2019, 2021) also faces problems. In particular, it could not explain the syntactic acceptability of (12) where, they say, we insist on making the two predicates in (8) be identical, hence leaving no space for meaning modulation.

(12) This table is red_1 and not red_2 and the property red_1 is identical to the property red_2

However, sentences concerning identity are not always interpreted as strict identities. In other words, they do not always entail that the pre-copular and the post-copular DPs are co-extensional (Fiorin & Delfitto 2024). Indeed, in a pragmatically supportive context, even possessive binding becomes acceptable:

(13) $Anne_i$ is identical to her_i mother

In (13) we do not take Anne and her mother to be co-extensional, but to share certain typical characteristics (e.g.: they behave in the exact same way, or they have a striking physical resemblance). We will argue that, similarly, (12) does not necessarily entail that red_1 and red_2 are co-extensional, therefore pragmatic considerations can still make the word-sequence non-contradictory.

Furthermore, we will argue that moving from logical identity to logical identification is the key to account for other examples, involving co-binding, considered problematic for language logicality, such as (14) and (15).

(14) $John_i$ is smarter than $himself_i$
(15) $John_i$ is $himself_i$

To conclude, non-typological explanations of (1)–(7) have not yet been decisively ruled out, and the typological explanation offered by Abrusán, Asher & Van de Cruys (2021) does not represent a defeat for language logicality but a redefinition of it: given that the peculiarity of (1)–(7) is due to types denoting a context-invariant logical meaning, logical considerations are relevant for syntactic formation after all.

References

Abrusán M., 2014, *Weak Island Semantics*, Oxford, Oxford University Press.

Abrusán M., N. Asher & T. Van de Cruys, 2021, «Grammaticality and Meaning Shift» in G. Sagi, J. Woods (eds.), *The Semantic Conception of Logic*, Cambridge (MA), Cambridge University Press, pp. 249–276.

Chierchia G., 2013, *Logic in Grammar: Polarity, Free Choice, and Intervention*, Oxford, Oxford University Press.

Chierchia G., 2021, «On Being Trivial: Grammar vs. Logic» in G. Sagi, J. Woods (eds.), *The Semantic Conception of Logic*, Cambridge (MA), Cambridge University Press, pp. 227–248.

Del Pinal G., 2019, «The Logicality of Language: A New Take on Triviality, "Ungrammaticality", and Logical Form», *Noûs*, 53, 4, pp. 785–818.

Del Pinal G., 2021, «The Logicality of Language: Contextualism vs. Semantic Minimalism», Manuscript, University of Illinois.

Fiorin G. & D. Delfitto, 2024, «Binding in copular sentences and the logic of identification», Presentation delivered at IGG49.

Fox D. & M. Hackl, 2006, «The Universal Density of Measurement», *Linguistics and Philosophy*, 29, 5, pp. 537–586.

Gajewski J., 2002, «L-analyticity and Natural Language», Manuscript, MIT.

Gajewski J., 2009, «L-triviality and Grammar», Handout, UConn Logic Group.

Pistoia-Reda S. & L. San Mauro, 2021, «On Logicality and Natural Logic», *Natural Language Semantics*, 29, 3, pp. 501–506.

Pistoia-Reda S. & U. Sauerland, 2021, «Analyticity and Modulation: Broadening the Rescale Perspective on Language Logicality», *International Review of Pragmatics*, 13, 1, pp. 1–13.

The Value of Conversation (and the Rise of Conversational Assistants)

Anna Drożdżowicz

Inland Norway University of Applied Sciences, Norway | anna.drozdzowicz@inn.no

Some conversations seem particularly valuable, some seem pointless. What is the value of having a conversation, in general? A first pass answer to this question is immediately suggested by a prominent tradition in philosophy of language: linguistic communication is a cooperative endeavor aimed at sharing information and influencing each other (e.g., Grice 1991; Lewis 2008). On this picture, having a conversation can be seen as valuable to the extent it allows for transferring information (i.e. acquiring true beliefs, knowledge, etc.) between interlocutors and coordinating on certain actions. This seems to capture an important way in which conversations are (instrumentally) valuable.

Arguably, the question about the value of conversation becomes much more pressing when we consider the growing presence and development of AI-powered speech and voice technologies. So-called intelligent conversational assistants such as Alexa or Siri have become popular commercial services (e.g. Turow 2021; McTear 2022), and are being developed to aid services in public sectors, such as healthcare and education. Recent developments in AI are likely to make such conversational agents even more powerful tools (e.g. Bansal et al. 2024; Gabriel et al. 2024). Ask your conversational assistant about how to order a taxi, you get your information. Arguably, you can learn from conversational assistants (cf. Mallory 2023). Thus, on the first pass answer, you can have a valuable conversation with them. And yet, one may still wonder whether and to what extent conversations with such assistants are valuable in the same way as those among regular interlocutors.

In this talk, I will focus on one important aspect of interactions with conversational assistants, i.e. that of illusion. Conversations with AI-powered conversational assistants may involve various 'local' illusions. My first goal will be to explicate what kind(s) of illusions may be involved, including, the illusion of: there being another interlocutor, there being a source of information provided, of interacting with that source and, in some cases, of having a relationship with them. On some views, our engagement with chatbots is best understood as a form of prop-oriented make-believe (Mallory 2023). Conversations with AI-powered assistants seem to be different from those among regular interlocutors with respect to how they relate to the truth. This makes our experience with them illusory (or virtual) in an interesting sense. But whether this makes such conversations less valuable (or dis-valuable) is an open and complicated question. My second goal will be to make some progress on it by discussing some classic thought experiments such as Nozick's experience machine, as well as recent arguments concerning entering virtual reality (e.g., Chalmers 2022).

Engineering Singular Concepts

Heimir Geirsson

Iowa State University, United States of America | geirsson@iastate.edu

Conceptual engineering is concerned with the evaluation and changes, hopefully for the better, of concepts, where concepts are thought of as representational devices. Included in

conceptual engineering are attempts to uncover the true meaning of a concept as well as attempts to find out what a concept should mean. The former bears the hallmarks of conceptual analysis and, as such, it doesn't seem to affect any changes in concepts. For example, while epistemologists have for centuries tried to specify the conditions for one having knowledge, the various accounts have tended to come up short when scrutinized. The concept *knowledge* remains somewhat elusive, and none of the attempts to clarify the meaning of 'knowledge' has changed the concept itself, the concept *knowledge*. The latter often involves providing an ameliorative or a revisionary meaning of a concept which, sometimes, arguably, resembles more the introduction of a new concept instead of refinement of the existing one. Well known examples are Sally Haslanger's amelioration of the concepts *woman* and *race* where she builds oppression into a definition of each concept. The discussion of conceptual engineering has focused on kind terms, and then kind terms for both natural kinds and socially constructed kinds. Singular concepts, concepts of individuals, have so far not been a part of the discussion. But singular concepts can be engineered and the nature of singular concepts and how they can be engineered can shed some light on the very nature of conceptual engineering. After presenting an account of singular concepts and some examples of their engineering, the lessons learned will be applied to two prominent accounts of conceptual engineering, namely the austerity framework of Herman Cappelen and the practical role account of Jennifer Nado. Neither account is satisfactory, and one way to show so is to bring singular conceptual engineering into the picture.

Singular concepts are concepts that we have about individuals. It has become common in recent years to talk about such concepts as mental files, and I will follow suit. Mental files are a tool that advocates of direct reference favor, as such files can play the role of Fregean senses without attaching semantic value to them. As a Millian I accept direct reference, believe that a proper name contributes its reference to the proposition expressed by the sentence it occurs in, and that truth-value is preserved when substituting co-referential names.

My account of files differs in important ways from that of Recanati. First, Recanati's framework does not talk about some elements in a file being more central than other. Second, Recanati relies on ER (epistemically relevant) relations, where those relations are causal relations between information in a file and the object the information is about. I avoid ER relations, one reason being that we can have files about objects that do not exist, such as Santa Claus. Santa cannot stand in a causal relationship to anything. Third, while Recanati states that files refer and that names borrow their reference from files, my account is more traditional in that I think that names refer. Names are one way of directing our attention to a particular file and placing information in that file and elicit information from that file. Fourth, while Recanati states, for example, in the prefix to *Mental Files in Flux* that "On my view, files only contain *predicative* elements."[i] I recognize that information and our representation of objects can consist of more than predicates.

So, how can one engineer singular concepts, i.e., how can one intentionally go about changing one's representations? There are several ways of doing so. I can, e.g., intentionally meet up with an old childhood friend that I have not had any contact with for decades. My mental file, singular concept, of the person will as a result receive significant updates. Or I can, depending on the nature of the information that I have, merge two files, or combine two concepts. The result will again be intentionally changing one's representation of the object the concepts represent. Or, as happened with the standard meter, I can update my information about the standard meter as the definitions of the meter have changed. Each

definition is more precise than the previous ones and each one leads to a more precise, and thus improved representation of the standard meter.

How does the information above affect our understanding of conceptual engineering? Cappelen presents a framework for conceptual engineering that emphasizes meaning revisions that result in a reference change. Nado presents a view that claims that the main target of conceptual engineering is classification procedures. The main goal of conceptual engineering, on her view, is to replace one classification procedure with another resulting, in most cases, in replacing one category with another. Both Cappelen and Nado present their accounts without relying on concepts. I follow the two in that regard; I don't focus on concepts as they are typically understood. Instead, my focus is on mental files, conceived as a repository of information. Mental files, on my view, do not have semantic properties and so do not refer. They are more properly thought of as conceptions that are realized as concrete particulars. The examples of engineering of singular concepts cannot be accounted for by the views presented by Cappelen and Nado. The engineering consists of modifying, updating the information in a given file. There is no meaning change that results in reference change, and there is no reclassification that takes place and hence no change in classification.

[i] viiiFrançois Recanati, *Mental Files in Flux* (Oxford: Oxford University Press, 2016).

Monday, 26/Aug/2024 17:10 – 17:50

Philosophy of Mind

Essential Representationalism about Perceptual Experience

Paweł Grad

University of Warsaw, Poland | pawelgrad@uw.edu.pl

In this paper, I explore the prospects for essential representationalism about perceptual experience (ER).

ER: For every phenomenal property in the class Φ, there is some representational property such that, necessarily, if perceptual experience has that phenomenal property then it has that representational property.

ER is a substantive philosophical claim because, if true, it can be used to explain the fundamental role of perceptual experience in grounding the intentionality of other mental states (Loar 2003; Kriegel 2011; Pautz 2013) and the cognitive-epistemic role of perceptual experience (Tye 2014; Dretske 2003). These explanatory virtues have led to the wide acceptance of ER among philosophers of perception. They have also, however, led to a general indifference towards pursuing independent arguments for ER. I contend that ER, in fact, requires an argument. This need is even greater after David Papineau's (2021) recent attack on ER.

The paper has the following structure. First I discuss methodological problems encountered by attempts to establish ER. Next, I defend ER by way of an imagination-based argument from visual emergence. Doing so, I hope, avoids the relevant methodological problems. I conclude by address some possible objections to my thesis and ER in general.

In the reminder of this abstract I restrict myself to briefly introducing my argument for ER presented in detail in the second part of the paper.

Take my current visual experience of a white coffee cup with blue spots (henceforth: Cup_E). Cup_E is a case of distinctively experiential presentation in the sense that it meets criteria specified by the following thesis:

Experiential Presentation: Even without interpretation or augmentation by higher-order mental states, perceptual experience represents its objects by rendering them phenomenally present to a subject's consciousness.

Phenomenal presence is sometimes cashed out in terms of transparency of experience vis-à-vis perceived objects and properties that are *not* represented as properties of the subject. But phenomenal presence is more than just transparency and objectivity of perceptual experience: perceptual experience by making things phenomenally present, seems to make us aware of the mind-independent particulars that are truth-makers of perceptual content (Pryor 2000: 547, fn. 37; Chudnoff 2012; Dorsch 2018; Berghofer 2020). I take the term "phenomenal presence" to be an intuitively understandable placeholder for the distinctively experiential and perceptual way of representing the environment.

Now, the question is whether the subject can have a non-representational phenomenal equivalent of Cup_E; call this equivalent $\Phi\text{-}Cup_E$. $\Phi\text{-}Cup_E$ would be phenomenally equivalent to Cup_E but devoid of its representational properties.

On CR (a negation of ER), the phenomenal characters of Cup_E and $\Phi\text{-}Cup_E$ consist of the same mental picture, only, in the former case, it happens to represent the blue-spotted white coffee cup. Now, suppose that a subject undergoes experience $\Phi\text{-}Cup_E$. However, at some point, this experiential episode changes in one aspect: $\Phi\text{-}Cup_E$ becomes Cup_E. This change will be a case of the emergence of the mental picture's representational aspect. CR does not provide reasons to think that this kind of representational-aspect emergence is impossible. In fact, CR predicts that such an emergence would remain phenomenology unaffected: the emergence of the representational aspect is phenomenally 'seamless'. I contend that this prediction is implausible in light of phenomenological considerations following from the *Experiential Presentation* thesis. An argument to this effect can be expressed in the following concise form:

The argument from visual emergence

P1 If ER is false, then the emergence of Cup_E from $\Phi\text{-}Cup_E$ does not change the phenomenal character constituted by the underlying mental picture.

P2 If the emergent experience is a case of experiential presentation, then the emergence of a representational aspect of experience in this case changes the phenomenal character constituted by the underlying mental picture.

P3 Cup_E is a case of experiential presentation.

P4 It is not true that the emergence of Cup_E from $\Phi\text{-}Cup_E$ does not change the phenomenal character constituted by the underlying mental picture.

∴ ER is true.

P1 is simply a CR's prediction. The conclusion follows from P1 and P4 by modus tollens. P4 follows from P2 and P3. P3 means that Cup_E is a case of perceptual experience in the sense implied by the *Experiential Presentation* thesis. *Experiential Presentation* is independently plausible and accepted by CRists (see Papineau 2021: 9–11, 95–97, 114–115). CRists must deny P2. As such, P2 is the only premise that requires support. I shall argue that our experience-based imagination strongly supports P2. CRist cannot reject P2 without undermining the *Experiential Presentation* thesis. CR has no resources to explain a genuinely experiential aspect of perceptual representation. Without augmentation by the *ad hoc* claims, CR predicts no phenomenal difference in cases of visual emergence in which we should expect such a difference (e.g. in Gestalt cases). Even if it does not conclusively rule out CR, the implausibility of this prediction speaks strongly in favour of ER.

References

Berghofer, P. (2020) 'Towards a phenomenological conception of experiential justification', *Synthese*, 197: 155–183.

Chudnoff, E. (2012) 'Presentational Phenomenology', in: S. Miguens & G. Preyer (eds.), *Consciousness and Subjectivity*. Frankfurt am Main: Ontos Verlag.

Dorsch, F. (2018) 'The Phenomenal Presence of Perceptual Reasons', in: F. Dorsch & F.Macpherson (eds.), *Phenomenal Presence*. Oxford: Oxford University Press.

Dretske, F. (2003) 'Experience as representation', *Philosophical Issues*, 13: 67–82.

Kriegel, U. (2011) *The Sources of Intentionality*. Oxford: Oxford University Press.

Loar, B. (2003) 'Phenomenal intentionality as the basis of mental content', in: M. Hahn & B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Cambridge, MA: MIT Press.

Papineau, D. (2021) *The Metaphysics of Sensory Experience*. Oxford: Oxford University Press.

Pautz, A. (2013) 'Does Phenomenology Ground Mental Content?', In U. Kriegel (ed.), *Phenomenal Intentionality*. Oxford: Oxford University Press.

Pryor, J. (2000) 'The skeptic and the dogmatist', *Noûs* 34: 517–549.

Tye, M. (2014) 'Transparency, qualia realism and representationalism', *Philosophical Studies*, 170: 39–57.

Monday, 26/Aug/2024 12:10 – 12:50
Philosophy of Language
Logic for Intransitive Coordination
Aidan Gray

University of Illinois Chicago, United States of America | grayaidand@gmail.com

Following recent usage, I'll use 'coordination' as a term for the relation that Fregeans have called 'sameness of sense'. For the Fregean, sameness of sense is a relation that can hold between representations of the same object. It has a special rational status: when coreferential representations share a sense, their coreference is rationally-relevant, in that it is a determinant of relations of entailment and inconsistency. For example, the Fregean holds that the difference between the subject who (rationally) believes that Dylan is a musician and that Zimmerman is not a musician, and the subject who (irrationally) believes that Dylan is a musician and that Dylan is not a musician, is that in the latter case the two representations of the relevant man are presented via the same sense, and thus their coreference is transparent to the subject.

To systematically characterize the way that the distribution of senses determines rational relations between representations, Fregeans (at least implicitly, and often explicitly) turn to standard first-order logic. They take the formal significance of *recurrence of constants* in a first-order language to model the rational significance of *recurrence of senses* in thought or language.

One benefit of using 'coordination' over 'sameness of sense' is that it allows us to explicitly raise the question of whether the relevant relation is *transitive* (whatever senses are, sameness of sense is a transitive relation). It is a substantive question whether, when the coreference of *a* and *b* is rationally-relevant, and the coreference of *b* and *c* is rationally-relevant, the coreference of *a* and *c* must also be rationally-relevant. Several recent discussions of coordination have discussed the possibility that coordination might be intransitive (see [1], [2], [3], [4], [5], [6]). But very little attention has been paid to the question of how we should understand the rational relations generated by intransitive coordination. This is a pressing question. If there is no coherent way to answer it, that would be decisive evidence against any theory that posits intransitive coordination. And there is reason to worry. As Fine (2007, pg. 119) writes:

"[...] the failure of transitivity [of coordination] has some radical implications for the conception of logical validity [. . .] the very idea of validity in virtue of logical form, as this is normally conceived, may also break down."

Without a response to this worry, we should not be confident that intransitive coordination is so much as intelligible.

In this paper, developing ideas by Fine, Pinillos, and Recanati, I show how the proponent of intransitive coordination should model the rational relations induced by intransitive coordination. That is to say, I develop a non-standard logical system that models the rational relations posited by the proponent of intransitive coordination.

I do this in three steps. First, I characterize the Fregean account of coordination in a more-than-usually-abstract way. In brief, we can think of the distribution of senses, for the Fregean, as *implementing* a more abstract idea: when two representations are coordinated there is a *semantic requirement* that they corefer (cf Fine 2007).

Second, I show that with only a small tweak, this can be altered to generate an intransitive picture of coordination: when two representations are coordinated there is a semantic requirement that *they do not have disjoint reference* (cf. Pinillos 2011, Recanati 2016).

Finally, I develop a logical system that models this kind of coordination. It turns out that any logic for intransitive coordination that satisfies certain very weak constraints has a surprising feature: if the logic is conservative with respect to standard first-order logic—in the sense that it gives the same verdict on arguments that involve only a transitive coordination-structure—it will have to employ an intransitive *consequence relation*. I describe a logical system, incorporating elements of free logic and three-valued logic, that satisfies this constraint. The logic deploys a consequence relation sometimes called "Strawson Entailment". A set of premises Strawson-entails a conclusion if and only if the truth of the premises guarantees the non-falsity of the conclusion. As is required to model intransitive coordination, Strawson entailment is an intransitive consequence relation. So the system allows that individual valid arguments do not always combine to make new valid arguments.

I close by outlining what this investigation reveals about the choice between the traditional, transitive, picture of coordination and this intransitive picture. There is a sense in which the traditional theorist loses nothing by accepting this new picture: where coordination is, in fact, transitive, the two accounts make the same claims about rational relations. The difference between them is that the intransitive account allows us to characterize the coordination structure in certain cases where classical rational relations break down.

References

[1] Contim, Filipe Drapeau Vieira. 2016. Mental Files and Non-Transitive De Jure Coreference. Review of Philosophy and Psychology, 7(2), 365–388.

[2] Fine, Kit. 2007. Semantic Relationism. Wiley-Blackwell.

[3] Goodsell, Thea. Is de jure coreference non-transitive?. Philosophical Studies 167 (2014): 291-312.

[4] Pinillos, N Ángel. 2011. Coreference and meaning. Philosophical Studies, 154(2), 301–324.

[5] Recanati, François. 2012. Mental files. Oxford University Press.

[6] Recanati, François. 2016. Mental files in flux. Oxford University Press

Should Language Models be Treated as Models? If so, of What?

Jumbly Grindrod

University of Reading, United Kingdom | j.grindrod@reading.ac.uk

In this talk, I am going to argue that language models (including large language models such as OpenAI's GPT series) should be treated as scientific models of external languages: languages understood as a social object and that are plausibly sets of linguistic conventions adopted by a community. In order to defend this view, I will first reject two related positions: the first claims that language models can be used as models of linguistic competence while the second claims that language models are (merely) models of their training data.

Many within computational linguistics, and specifically the field of distributional semantics, are excited about the prospect of language model technology as a new form of scientific inquiry into language (Baroni, 2022; Lenci, 2008; Sahlgren, 2008; Westera & Boleda, 2019). Perhaps the most vocal recent proponent of this view is (Piantadosi, 2023), who claims that language model technology challenges some of the core claims of the generative linguistic tradition. Although they don't phrase it this way, the best interpretation of this view is that language models can be treated as *models* of linguistic competence, and so we can then inspect the model as a way of investigating linguistic competence. But the idea that deep learning neural networks could inform linguistic inquiry in this way has been criticized by Chomsky (Chomsky et al., 2023; Norvig, 2012) and others (Dupre, 2021; Veres, 2022). Roughly put, these critics worry that language models are blank slate systems (i.e. they do not have the same innate restrictions as humans) that simulate speaker performance without emulating speaker competence. Although there is evidence being produced from the emerging probing classifier literature that attenuates the strength of the points made by these critics, I will argue that they are nevertheless right.

Many who are skeptical of the possibility of language models providing linguistic insight have instead claimed that language models are merely models of their training data. After all, language models are constructed by setting a language model the task of predicting new text given what has come previously according to the distributional properties of the data it was trained on. This is the second position I will consider. A similar view can be found in Chiang's (2023) suggestion that language models are best thought of as compressions of their training data, as a jpeg is of a higher resolution image. This view also has an affinity with Kilgarriff's (1997) famous claim that word meanings only exist relative to the statistical properties of corpora. However, I will argue against this position, for in considering the success of a language model we do not evaluate its success in the language prediction task it is trained on, but set the model to work on new evaluation tasks. One of the amazing insights of language model technology is that these systems are able to perform so well across a wide range of natural language processing tasks. The nature of the evaluation tasks for such models – as well as their success in them – reveals that we are not holding models to a standard internal to the training corpus but are instead testing the extent to which they track something consistent across both their training set and evaluation sets.

I will argue then that language models should be thought of as models of the external language understood as a social object: the *E-language* in Chomsky's (1986) terms. What language models are trained on is the actual activity of a language, where all instances across training and evaluation sets are taken to share the feature of being part of the wider language. Viewed through this lens, we are able to see the exciting possibility that language models bring, for they provide us with a way of exploring E-languages that was not available

before. If E-languages are a set of social conventions, then they are undoubtedly highly complex objects, and if we acknowledge the fact that any speaker's cognizance of that set of conventions is going to be incomplete and imperfect, then access to that complex object has previously looked fraught with difficulty. This is partly why Chomsky (1986) has taken there to be no point in positing E-languages. But now that we are able to construct models of an E-language, and in doing so bypass the cognitive domain in a way that wasn't possible before, we have a new and exciting way of investigating them. I will finish by drawing upon recent work in philosophy of science on the use of deep learning models in scientific practice in order to further support the positive view defended here (Creel, 2020; Shech & Tamir, 2023; Sullivan, 2022, 2023).

References

Baroni, M. (2022). *On the proper role of linguistically-oriented deep net analysis in linguistic theorizing* (arXiv:2106.08694). arXiv. https://doi.org/10.48550/arXiv.2106.08694

Chiang, T. (2023). Chatgpt is a blurry JPEG of the web. *The New Yorker*. https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web

Chomsky, N. (1986). *Knowledge of Language: Its nature, origin, and use*. Praeger.

Chomsky, N., Roberts, I., & Watumull, J. (2023, March 8). The false promise of chatgpt. *The New York Times*. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Creel, K. A. (2020). Transparency in Complex Computational Systems. *Philosophy of Science*, *87*(4), 568–589. https://doi.org/10.1086/709729

Dupre, G. (2021). (What) can deep learning contribute to theoretical linguistics? *Minds and Machines*, *31*(4), 617–635. https://doi.org/10.1007/s11023-021-09571-w

Kilgarriff, A. (1997). I don't believe in word senses. *Computers and the Humanities*, *31*(2), 91–113. https://doi.org/10.1023/A:1000583911091

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, *20*(1), 32.

Norvig, P. (2012). Colorless green ideas learn furiously: Chomsky and the two cultures of statistical learning. *Significance*, *9*(4), 30–33. https://doi.org/10.1111/j.1740-9713.2012.00590.x

Piantadosi, S. (2023). *Modern language models refute Chomsky's approach to language*. LingBuzz. https://lingbuzz.net/lingbuzz/007180

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics*, *20.1*, 33–53.

Shech, E., & Tamir, M. (2023). *Understanding from Deep Learning Models in Context* [Preprint]. https://philsci-archive.pitt.edu/21296/

Sullivan, E. (2022). Understanding from Machine Learning Models. *The British Journal for the Philosophy of Science*, *73*(1), 109–133. https://doi.org/10.1093/bjps/axz035

Sullivan, E. (2023). Do Machine Learning Models Represent Their Targets? *Philosophy of Science*, 1–11. Cambridge Core. https://doi.org/10.1017/psa.2023.151

Veres, C. (2022). Large language models are not models of natural language: They are corpus models. *IEEE Access*, *10*(Journal Article), 61970–61979. https://doi.org/10.1109/ACCESS.2022.3182505

Westera, M., & Boleda, G. (2019). Don't blame distributional semantics if it can't do entailment. *IWCS*. https://doi.org/10.18653/v1/W19-0410

Understanding Oneself
J.P. Grodniewicz

Jagiellonian University, Poland | j.grodniewicz@gmail.com

§ 0.

What does it take to understand oneself? I argue that it involves grasping coherence-making relations between constituents of a body of information about oneself. As such, self-understanding is best characterized as a kind of holistic understanding of a subject domain, often called *objectual understanding*. What is special about self-understanding is the domain — oneself. This characterization enables us to explain what it means to deepen one's self-understanding and why self-understanding can neither be reduced to self-knowledge nor self-narrative.

The paper is divided into three parts.

§ 1.

I start by analyzing which of the kinds of understanding commonly discussed in contemporary epistemology fits best for an account of self-understanding. After rejecting the views that self-understanding is a kind of *understanding-that*, *understanding-why*, or *understanding-how*, I argue that we should characterize self-understanding as a kind of *objectual understanding* (for influential accounts of objectual understanding see, e.g., Kvanvig, 2003; Riggs, 2003, 2009; Grimm, 2011; Elgin, 2017). To wit, self-understanding is more like understanding marine biology than understanding *why* humpback whales migrate annually or *how* to stimulate coral reef restoration.

Given that objectual understanding is commonly characterized as consisting of grasping relations between constituents of a body of information, characterizing objectual understanding of any subject matter — oneself not being an exception — requires answering two questions: "What do we mean by grasping?" and "What is the relevant information?"

To answer the first question, I follow many contemporary epistemologists in claiming that grasping is a type of ability or a set of abilities (e.g., Wilkinfeld, 2013; Hills, 2016; Elgin, 2017). On my account, understanding oneself essentially involves possession of an ability (or a set of abilities) to represent and manipulate a body of information about oneself.

But what is the relevant information? Here, I appeal to the psychological notion of "self-concept," understood as "a mental representation…what we bring to mind when we think about ourselves" (Neisser, 1997, p. 3), and argue that self-understanding is grasping coherence-making relations between constituents of one's self-concept.

§ 2.

The view outlined above allows us to offer a fine-grained account of what it means to deepen one's self-understanding and explain why self-understanding is not reducible to self-knowledge.

Almost all we do, from simply living our lives and letting things happen to us, through reading novels and engaging with art, to interacting with other people including our friends and therapists, may lead to deepening our self-understanding. I argue that it happens along three main dimensions. We can increase: (1) the *richness* of our self-understanding by increasing the amount and scope of information we possess about ourselves; (2) *factivity* by increasing the truths-to-falsehood ratio in the body of information, and (3) *coherence* by recognizing more connections between the pieces of information we have about ourselves.

Bits of self-knowledge contribute to the body of information whose grasping constitutes self-understanding. Therefore, knowing more about oneself typically results in deeper (richer and more factive) self-understanding. At the same time, self-understanding has important features that distinguish it from knowledge. It is compatible with having at least some false information about oneself and essentially consists of an ability to grasp coherence making relations between relevant pieces of information. Moreover, we typically do not consider knowledge to be gradable.

Given that non-factivity (e.g., Riggs, 2009; Elgin, 2017), gradeability (e.g., Zagzebski, 2001; Kvanvig, 2003; Riggs, 2003; Pritchard, 2009), and the essential role of grasping (Kvanvig, 2003; Riggs, 2003; Grimm, 2014; Elgin, 2017) are usually enumerated as what makes objectual understanding irreducible to knowledge, I conclude that self-understanding is irreducible to self-knowledge.

§ 3.

In the last part of the paper, I argue that while building a self-narrative is a way of developing self-understanding, it is not the only way.

Narratives afford (some level of) understanding of their protagonists by establishing coherence-making relations between their life events; the protagonist of a self-narrative is oneself; thus, a self-narrative can afford (some level of) self-understanding.

Nevertheless, as argued, by Strawson (2004), many people think about themselves in a predominantly episodic fashion, in which case "one does not figure oneself, considered as a self, as something that was there in the (further) past and will be there in the (further) future" (2004, p. 430). The view I defend enables us to explain why both predominantly episodic and predominantly diachronic individuals can attain a certain level of self-understanding and further deepen it. To understand oneself is to grasp coherence making relations between constituents of a body of information about oneself. Predominantly diachronic individuals tend to do it through self-narrative; while predominantly episodic individuals are less inclined to use this particular tool.

In short, there are more ways of grasping relations between elements of self-concept than just by putting them into a narrative. Thus, not all self-understanding is narrative self-understanding.

References

Elgin, C. Z. (2017). *True Enough*. Cambridge, MA: MIT Press.

Grimm, S. (2011). Understanding. In D. Pritchard & S. Bernecker (Eds.), *The Routledge Companion to Epistemology*. Routledge.

Grimm, S. (2014). Understanding as knowledge of causes. In *Virtue epistemology naturalized* (pp. 329–345). Springer.

Hills, A. (2016). Understanding Why. *Noûs*, *50*(4), 661–688.

Kvanvig, J. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press.

Neisser, U. (1997). Concepts and self-concepts. In U. Neisser & D. A. Jopling (Eds.), *The conceptual self in context: Culture, experience, self-understanding* (pp. 3–12). Cambridge University Press.

Pritchard, D. (2009). Knowledge, Understanding and Epistemic Value. *Royal Institute of Philosophy Supplement*, *64*, 19–43

Riggs, W. D. (2003). Understanding "Virtue" and the Virtue of Understanding. In M. DePaul & L. Zagzebski (Eds.), Intellectual Virtue:34 Perspectives From Ethics and Epistemology (pp. 203–226). Oxford: Oxford University Press.

Riggs, W. D. (2009). Understanding, Knowledge, and the Meno Requirement. In A. Haddock, A. Millar, & D. Pritchard (Eds.), *Epistemic Value*. Oxford: Oxford University Press.

Strawson, G. (2004). Against Narrativity. *Ratio*, *17*(4), 428–452. https://doi.org/10.1111/j.1467-9329.2004.00264.x

Wilkenfeld, D. A. (2013). Understanding as representation manipulability. *Synthese*, *190*(6), 997–1016. https://doi.org/10.1007/s11229-011-0055-x

Zagzebski, L. (2001). Recovering Understanding. In M. Steup (Ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Oxford University Press.

Measuring Conceptual Inflation: The Case of "Racist"

Nat Hansen[1], Shen-yi Liao[2]

[1]University of Reading (UK); [2]University of Puget Sound (USA) | n.d.hansen@reading.ac.uk

Is the term 'racist' being applied so widely that it is losing its moral force? Theorists and pundits from across the political spectrum think that it is. They call such a change of meaning 'conceptual inflation' and argue that we should try to stop it by restricting the use of 'racist'. But is there any evidence that 'racist' is becoming conceptually inflated? Economists do not track currency inflation with mere vibes; they use measurements such as the consumer price index. For example, using CPI, it is observable that there has been considerable US dollar inflation over the last three decades or so: $1.00 in 1989 has the purchasing power of $2.47 in 2023. Has 'racist' undergone a similar transformation, as critics of conceptual inflation contend? How could that be measured? Before plunging into the normative debate, we should gather systematic empirical evidence about whether the purported conceptual inflation of 'racist' is in fact happening. That is what we do in this paper: we use an apparent time study, a technique drawn from sociolinguistics (Bailey 2004), to measure whether 'racist' is undergoing conceptual inflation. An apparent time study looks at the linguistic behavior of people of different ages as a way of measuring linguistic variation over time. We make a distinction between two processes that constitute conceptual inflation: *extension expansion* and *intensity bleaching*. Across two pre-registered studies (n = 417 and n = 395), we find equivocal evidence that 'racist' has undergone extension expansion—its reference class may have slightly expanded—but no evidence that it has undergone intensity bleaching—meaning that its moral force remains the same. When we compare 'racist' with other terms that we have reason to think have undergone meaning change, like 'queer', we find that the conceptual inflation of 'racist', as it is standardly understood, is an illusion.

Blum (2002a, 31) gives a clear statement of the worry about the conceptual inflation of 'racism' and 'racist':

To name an act or a person "racism" or "racist" is particularly severe condemnation. But the terms are in danger of losing their moral force, for they have been subject to conceptual inflation (overexpansive usage) and moral overload (covering morally too diverse phenomena), thus inhibiting honest interracial exchange. (Blum 2002a: 31)

Blum also makes a causal claim that overuse of an expression leads to diminished moral force:

That overuse in turn feeds a diminishing of "racism"'s moral force, and thus contributes to weakened concern about racism and other racial ills. (Blum 2002b: 206)

For our experiments, we devised the following two questions to operationalize the two aspects of the purportedly changing use of "racist" that Blum identifies:

1. *Extension*: What % of people do you think can reasonably be called 'racist'?
2. *Intensity* (of moral criticism): How bad is it for a person to be called 'racist'?

If Blum is right, an apparent time study should show (a) an increase in the *extension* of 'racist' (that is, younger people should apply it to a higher percentage of the population)—what we call *extensional expansion*—and (b) a decline in the *intensity* of 'racist' (that is, younger people

should find its use less bad than older people, because it now covers more diverse phenomena than it used to)—what we call *intensity bleaching*. If his causal claim is correct, we should at least see a (negative) correlation between larger extension and lower intensity associated with 'racist'.

Participants responded to prompts about the extension and intensity of 'racist' and other expressions, including 'queer'.

We found that while the extension of 'racist' was slightly greater for young people than for old people, we did not find any difference in the intensity of 'racist' between the age groups. That contrasts with what we found with the expression 'queer', which showed differences in both extension and intensity between young people and old people.

Our results show that the apparent time method (comparing younger and older language users) is sensitive enough to detect meaning change, given the assumption that 'queer' has undergone some change in meaning over the past three decades. Moreover, while there has been extension expansion with 'racist', this change is smaller in magnitude compared to the extension expansion with 'queer'. At the same time, while we found no evidence for intensity bleaching with 'racist', there has been intensity bleaching with 'queer', which is likely an effect from intentional reclamation efforts.

While we did find evidence that 'racist' has undergone some extension expansion (over 60s applied it more restrictively than the under 30s), we did not find any evidence in our two studies that 'racist' has either of the other two features that are typically combined in worries about conceptual inflation: intensity bleaching and an inverse correlation between extension and intensity. If the conceptual inflation of 'racist' is as rampant and problematic as many critics have proposed, our results should be surprising. We conclude by comparing worries about conceptual inflation to worries about moral and linguistic decline and argue that the conceptual inflation of 'racist', as it is standardly understood, is an illusion.

References

Bailey, Guy (2004). "Real and Apparent Time", in J.K. Chambers and Peter Trudgill and Natalie Schelling-Estes (eds.), *The Handbook of Language Variation and Change*, Oxford: Blackwell.

Blum, Lawrence (2002) *'I'm Not a Racist, But...': The Moral Quandaries of Race*. Ithaca, NY: Cornell University Press.

Wednesday, 28/Aug/2024 12:10 – 12:50

Philosophy of Language

Testing Name Swapping: Is Beyoncé Really Famous?

Jussi Haukioja, Jeske Toorman

NTNU Trondheim, Norway | jussi.haukioja@ntnu.no

Since Machery et al. (2004), a number of experimental studies have been conducted on the reference of proper names. Even though details vary, all existing studies have featured vignettes with a structure that models Kripke's famous Gödel and Jonah cases, followed by a question probing the test subjects' reference assignments, where one of the two options has been interpreted as supporting the causal-historical account, and the other as supporting descriptivism (e.g. Devitt & Porot 2018, Li et al. 2018, Domaneschi & Vignolo 2020). These studies have, for the most part, claimed to find evidence for the causal-historical account, and against descriptivism. Doubts remain, however, about the validity of the setup. In particular (cf. Author XXXX), it is reasonable to expect that the test subjects will associate information stated in the vignette with the name in question, and that the response alternative that has been interpreted as supporting the causal-historical account is consequently fully compatible with a descriptivist theory of names. If so, these responses

do not support the causal-historical account, after all. In earlier work (Author & Author XXXX) we have substantiated this worry by showing that similar proportions of purportedly causal-historical responses can be elicited with analogous vignettes featuring terms of which the causal-historical account is very likely false.

In this paper, we present results from a new set of experiments. Our setup is methodologically novel in two ways. First, unlike in earlier work, the vignettes we used did not use the proper names the reference of which is under study, but merely mentioned them. This allowed us to remove what we took to be problematic aspects of the setups modelled on Kripke's Gödel and Jonah cases. Inspired by Johnson & Nado (2014), our vignettes described two individuals, but referred to them as "X" and "Y", instead of the proper names of interest. In the vignettes, X is described as possessing the properties typically associated with name N, while Y is a different individual. However, the vignettes state that Y, and not X, was originally called by name N, while X was called by another name. The vignettes then go on to state that the two names got mixed up at some point, which is why speakers today typically associate the properties they do with name N. The subjects were then asked to judge which individual, X or Y, is the referent of N. The causal-historical account predicts that name N continues to refer to Y (the original bearer of N), while the descriptivist prediction is that name N, in current use, refers to X (who satisfies all the properties associated with N). These vignettes featured familiar names such as "Isaac Newton" and "Beyoncé".

The second novelty of our setup is that, in addition to the vignettes just described, we used fully analogous vignettes which featured, instead of names, terms of which the causal-historical account is plausibly false, namely artefact terms (such as "guitar" and "shovel") and definitional terms (such as "bachelor" and "prime number"). This was done to test the validity of the setup. It is prima facie plausible to suppose that the extensions of artefact terms and definitional terms are determined by associated intended function, and associated descriptions or definitions, respectively. If the pattern of responses is found to be significantly different for these terms, as compared to proper names, such that subjects prefer the opposite response in an otherwise analogous setting, this would show that the participants' responses are tracking a difference in how the relevant terms' extensions are determined, and are thus not the result of some feature of the setup that has nothing to do with which theory of reference is true. If so, the responses predicted by the causal-historical account for the vignettes featuring names would provide much stronger support for the theory than previous studies have done. This was exactly what we found. For the proper name cases, a clear majority of responses (~64%) were as predicted by the causal-historical account. For the artefact terms, by contrast, a very clear majority of responses (~74%) were as predicted by the descriptivist account, and for the definitional terms, this majority was even greater (~81%). Our first conclusion is that our setup represents a major step forward methodologically, in experimentally testing the reference of proper names. Our second conclusion is that the reference of proper names is unlikely to be determined in the same way as the extensions of artefact terms and definitional terms. Our results thereby provide tentative empirical support for the causal-historical account of names.

References

Devitt, M. & Porot, N. (2018). The Reference of Proper Names: Testing Usage and Intuitions. Cognitive Science, 42(5), 1552–1585.

Domaneschi, F., & Vignolo, M. (2020). Reference and the Ambiguity of Truth-Value Judgments. Mind & Language, 35(4), 440–455.

Johnson, M. & Nado, J. (2014). Moderate Intuitionism: A Metasemantic Account. In Booth, A. & Rowbottom, D.

(Eds.). Intuitions (pp. 68-90). Oxford: Oxford University Press.

Li, Liu, L., Chalmers, E., & Snedeker, J. (2018). What Is in a Name? Cognition, 171, 108–111.

Machery, E., Mallon, R., Nichols, S. & Stich, S. P. (2004). Semantics, Cross-Cultural Style. Cognition, 92(3), B1–B12.

Can Original Content Determine Phenomenal Character?

Amir Horowitz
Open University of Israel, Israel | amirho@openu.ac.il

Representationalism about phenomenal consciousness is the view that phenomenal character is determined by representational (i.e., intentional) content. According to pure representationalism, phenomenal character is wholly determined by intentional content, whereas according to impure representationalism, it is determined by intentional content plus something else, usually some functional-cognitive property. Most representationalists endorse the impure version, plausibly because there are non-phenomenal representations and it isn't easy to distinguish in representational terms between them and phenomenal representations. But there are exceptions. Thus, according to Bourget (2010), phenomenal character is nothing but content – it is un-derived (otherwise put – original) content. Bourget's view appears to account successfully for the distinction between phenomenal and non-phenomenal representations, since (or so most philosophers believe) non-phenomenal representations do not possess their contents originally, but only derivatively: their contents are derived from the contents of phenomenal states with which they maintain some connections. (Another proponent of pure representationalism is Thau 2002. For impure representationalism see, e.g., Tye 1995, Dretske 1995, and Jackson 2004.)

The main purpose of this talk is to undermine the view that phenomenal consciousness is determined by original content. My claim is that either this view fails to distinguish between phenomenal and non-phenomenal representations, or it involves a non-representational element, and so is in fact an impure representationalist view. A subsidiary goal of the paper is to draw a distinction between two kinds of phenomenal representationalism on a dimension different from that of pure versus impure distinction – a subjectivist kind and a deflationist kind. This distinction is important for its own sake, but it will also play a role in the suggested case against Bourget's view.

The issue of original versus derived intentionality concerns the source of the contents of entities, rather than the identity of these contents. A conscious mental state (which is assumed to have its content originally) and a (mental or non-mental) non-conscious state or act (which is assumed to have its content derivatively) can share content: to ascribe to them the content that p originally and derivatively, respectively, is to ascribe to them not only different properties but also the same property, namely the same intentional content. So, I will argue, if it is the content of a conscious mental state that is supposed to fully determine its phenomenal character on the view under consideration, then according to this view the conscious state and the non-conscious one must have the same phenomenal character and (a fortiori) the non-conscious state comes out conscious. It might be thought that this predicament can be avoided in a reductive framework of intentionality. In such a framework, the determiner of a state's possession of original content is different from the determiner of another state's possession of the semantically identical derived content, and we can take the phenomenality of a conscious state to be constituted by the determiner of this state's possession of its content rather than by its content itself. This emendation of

representationalism seems to reserve its spirit, and it well coheres with an important motivation for it, which is to ultimately ground phenomenality in the "natural". However, I will argue that even if we can motivate this view as a theory of phenomenality, what enables it to distinguish between phenomenally conscious representational states and non-conscious ones is precisely the fact that the determiner of the possession of original content according to it is a determiner of the specific way in which the bearer of original content possesses it, rather than of its mere possession of content. So this view is a version of impure reprsentationalism.

But perhaps representationalists can have a way out. We should distinguish between two sorts of views according to which intentionality determines phenomenality, one that takes phenomenality to be thin and another that takes intentionality to be thick. On the latter view, intentionality is "shot through with subjectivity", as McGinn (1991) puts it (and, thus, thick), and its subjective nature enables it to determine the subjective nature that is associated with phenomenally conscious states. (Searle (1992) also seems to endorse such a subjectivist phenomenal representationalism.) On the former view, intentionality is laden with nothing but its semantic content, understood as sheer aboutness, and so phenomenality consists of nothing more (and is, then, thin). (Pure) representationalism thus understood may be said to embody a deflationist view of phenomenality, one which takes subjectivity out of consciousness or downplays it. Most current representationalist views do not take intentionality to be "shot with subjectivity", and so may be said to be deflationists. But it might seem that "original-content- representationalism" such as Bourget's can incorporate the subjectivist version and thus single out phenomenal representations. The idea might be that original content and only original content is shot with subjectivity, and it thus endows representations that possess it with phenomenality. However, we shall see that accounts of phenomenal consciousness in terms of intentionality that is shot with subjectivity are doomed to fail on pain of circularity – their explanations of what makes something phenomenal are virtus dormitiva explanations. Trivially, original-content-representationalism inherits this flaw.

Lying with Presuppositions — An Anaphoric Account
Antonina Jamrozik
University of Warsaw, Poland | a.jamrozik@uw.edu.pl

In my talk, I want to tackle the idea of lying with presupposition from a previously unexplored angle. The idea that one can lie with presuppositions is best spelled out in Viebahn 2020. There, he proposes a commitement-based view of lying, contrasted to saying-based definitions of lying (see eg. Stokke 2018, Saul 2012). According to Viebahn, a lie occurs when the content of a communicative act performed is the proposition p, the person performing this act commits herself to p and believes it to be false. Such definition of lying makes no reference to the concept of assertion, encompassing the cases where one lies with presupposition.

However, as it was pointed out by Gaszczyk (2023), Viebahn does not want to abandon the traditional definition of lying — he emphasises that in all the cases analysed by him in his paper the communicative act performed by the speaker is asserting, and goes as far as to claim that the committal communicative act in his definition can be understood as asserting. This is a consequence that Gaszczyk disagrees with and finds unnecessarily

conservative, as well as damaging to the distinction between assertion and presupposition. What Gaszczyk proposes is to get rid of Viebahn's claim that uninformative presuppositions cannot fulfil his definition of lying (because one cannot commit to something that is already in the common ground) and suggests that it is indeed possible to lie with uninformative presuppositions. While I agree with Gaszczyk with regards to his criticism of Viebahn's commitment to equating lying with assertion, I think his claim that one can lie with uninformative presuppositions is unsubstantiated and counterintuitive.

In my talk, I want to propose an approach that allows one to get rid of the idea that it is only possible to lie using assertions (even if the lie is due to a presupposition present in a given assertion) without committing oneself to the counterintuitive consequence that it is possible to lie with uninformative presuppositions. This approach is based on the idea that presuppositions should be treated as anaphors, albeit having more descriptive content (Van der Sandt 1992). This difference allows them to establish the discourse referent if there isn't one already present in the common ground. If there is one, presuppositions act just like anaphors — they get their content from the information previously established in the discourse. While Van der Sandt's goal was to explain presupposition projection as anaphora resolution, his idea can be used to explain why it is possible to lie with informative presuppositions, and not with uninformative presuppositions, without appealing to the notion of assertion. In the approach I propose this is explained by appealing to the anaphoric and descriptive properties of presuppositions. If the presupposition is considered informative, it means that there has not been enough information in the common ground, and this is where its descriptive content comes into play — it creates an antecedent for it in the common ground. If the proposition is uninformative, its antecedent is already represented in the common ground, and therefore the speaker's commitment to it is minimised. Such approach gets rid of the problematic qualities of both Viebahn's and Gaszczyk's accounts, while preserving their good parts — relying on commitment and allowing for lying with presuppositions even in conversational acts that are not assertions, such as questions.

The plan of the talk is the following: I start with a brief overview of Viebahn's account, as well as Gaszczyk's critique of it, working on examples from their papers. Then, I give an overview of Van der Sandt idea that presuppositions are anaphors, focusing on the question of how it would work in the case of uninformative presuppositions. Finally, coming back to Viebahn and Gaszczyk, I show how the problems in their accounts can be resolved using Van der Sandt idea.

References:

Gaszczyk, G. (2022). Lying with uninformative speech acts. Canadian Journal of Philosophy, 52(7), 746-760.

Saul, J. M. 2012: Lying, Misleading, & What is Said. Oxford: Oxford University Press.

Stokke, Andreas. 2018. Lying and Insincerity. Oxford: Oxford University Press.

Van der Sandt, R. A. (1992). Presupposition projection as anaphora resolution. Journal of semantics, 9(4), 333-377.

Viebahn, Emanuel. 2020. "Lying with Presuppositions." Noûs 54 (3): 731–51.

The Functional Composition of Thoughts

Alexander Johnstone

University of Pittsburgh, United States of America | alj100@pitt.edu

Frege scholars have long wondered how the thought expressed by a sentence relates to the senses of that sentence's constituent expressions. According to the orthodox answer, a thought is a whole which contains, as its constituent parts, the senses of the relevant expressions (Dummett 1973; Heck and May, 2011). Others claim that thoughts compose functionally: senses of functional expressions, such as predicates, are themselves functions. On this heterodox approach, there is no need to hold that a thought contains the relevant senses as its parts (Church, 1951; Pickel, 2021).

In this paper, I will present a novel argument for the functional approach. However, I do not want to engage in Frege interpretation. I want to engage instead with the many contemporary philosophers of language and mind who help themselves to the Fregean notion of a thought. Indeed, my argument is, in one respect, crucially different from the arguments put forward so far by Frege-interpreters favourable to the functional approach. These arguments are presented as turning on an idiosyncratic commitment, one which Frege just so happened to hold (*reference to senses*). By contrast, the argument I produce in this paper begins upstream from there: it is not dependent on any commitment that one who goes in for the notion of a thought can reasonably opt out of.

I first argue that anyone who takes the notion of a thought seriously must accept *reference to thoughts.* This is (with respect to English) the claim that one can *refer* to a thought by prefixing "that" to a sentence which ordinarily expresses the relevant thought.[1] One who goes in for the notion of a thought must thus accept that a that-clause can be used to refer to the thought which the embedded sentence ordinarily expresses. The paper then contends that *reference to thoughts*, together with three additional but innocuous claims, forces the functional approach to thought composition.

The first additional claim, *functional compositionality of reference*, is this: the referents (semantic values) of the names into which a complex expression can be logically analysed yield the referent of the entire complex expression through functional application. The semantic value of a complex whole expression is thus the result of applying a function which is the value of some component expression to the value of some other component expression. In formal semantics, the principle has obtained the status of a virtually unchallenged orthodoxy (Heim and Kratzer, 1998). The second claim is this: a sentence as it appears inside a that-clause has the same (syntactic and semantic) structure as when it occurs alone. One who has mastered the semantic and syntactic structure of "Snow is white" exercises her capacity to discern that very same structure in understanding "S believes that snow is white". Following Pickel, I call this *embedded structure* (2021). Finally, the third claim: *transparency*. *Transparency* states that a that-clause which refers to a thought cannot be understood unless one thinks the very thought it names. This has been noticed before by a variety of authors (i.e. Peacocke 2008). One cannot understand "It has not yet been proven that every even integer greater than two is the sum of two prime numbers" but fail to understand "Every even integer greater than two is the sum of two prime numbers". The sense *that every even integer greater than two is the sum of two prime numbers* is thus transparent, for grasping this sense involves grasping the very thought it determines, namely *every even integer greater than two is the sum of two prime numbers.*

These are all the premises needed for the argument. I first show that *reference to thoughts* and *transparency* together yield *reference to senses*. (*Reference to senses* says that an expression inside a that-clause refers to the sense it ordinarily expresses. Where *reference to thoughts* claims that an entire that-clause refers to a thought, *reference to senses* governs the constituent expressions of a that-clause.) Then, relying on Pickel's 2021, I argue that *reference to senses*, *functional compositionality of reference*, and *embedded structure* together yield the functional approach to thought composition.

The dialectical significance of this new argument is as follows. In the context of Frege interpretation, arguments for the functional approach to thought composition have standardly begun with *reference to senses*. But *reference to senses* has the look of an optional commitment, one which Frege could have perhaps done without. This has made it seem as if one could take the notion of a thought on board without accepting that thoughts compose functionally, so long as one does not follow Frege in accepting *reference to senses*. Indeed, this is precisely Dummett's reaction: he "charitably" amends Frege by rejecting *reference to senses*. My claim that *reference to senses* in fact follows from *reference to thoughts* modifies the dialectical situation. For no one who goes in for the notion of a thought should deny either *reference to thoughts* or *transparency*. (Indeed, Dummett does not). This shows that one cannot in fact go in for the notion of a thought without committing oneself to *reference to senses*. And this, in turn, forces the functional approach.

Bibliography

Church, Alonzo (1951). "A Formulation of the Logic of Sense and Denotation" in P. Henle, H.M. Kallen & S. K. Langer (eds.), *Structure, Method and Meaning* (New York: Liberal Arts Press), pp. 3 – 24.

Dummett, Michael (1973). *Frege: Philosophy of Language* (London: Duckworth).

Evans, Gareth (1982). *Varieties of Reference* (Oxford: Oxford University Press).

Heck, Richard, & May, Robert (2011). "The Composition of Thoughts", *Noûs* 45: 126–66.

Heim, Irene, & Kratzer, Angelika (1998). *Semantics in Generative Grammar* (Malden, MA: Blackwell Publishers).

Künne, Wolfgang (2015). "Frege on That-Clauses" in B. Weiss (ed.), *Dummett on Analytic Philosophy* (London: Palgrave Macmillan), pp. 135 – 173.

McDowell, John (1996). *Mind and World* (Cambridge: Harvard University Press).

Peacocke, Christopher (2008). *Truly Understood* (New York: Oxford University Press).

Pickel, Bryan (2021). "The Functional Composition of Sense", *Synthese* 199: 6917-6942.

[1] Different languages will of course admit of different versions of the principles.

Monday, 26/Aug/2024 17:10 – 17:50
Philosophy of Language
Negotiating Scripts with Normative Generics

Hadil Karawani[1], Daniel Skibra[2]
[1]University of Konstanz, Germany; [2]University of Konstanz, Germany | danielskibra@gmail.com

Normative generics and discourse
The distinction between generic sentences ("generics") and non-generic sentences is by now well known: generics express generalizations without any overt characterization of the distribution of the kind members with respect to the property being ascribed (in contrast to overt quantifier expressions, for example). But theorists have also posited a distinction within the category of generics themselves. Descriptive generics merely describe the above mentioned generalizations. The second kind of generic, which philosophers have recently started to call "normative generics", are thought to have a distinctly normative flavor to the

kinds of generalizations they express. (Cf. Burton-Roberts 1976, 1977; Cohen 2001; Leslie 2015) Even if the distinction is a bit vague, the contrast between descriptive and normative generics turns out to be quite striking, as (1) and (2) show. The (a)-sentences provide examples of descriptive generics, describing putative regularities in the world, whereas the (b)-sentences seem to make something closer to a normative claim.

(1)   a. Rugby players don't cry.
        b. Boys don't cry.
(2)   a. Philosophers work at universities
        b. Philosophers care about the truth.

Given the descriptive sense of (2a), arguably not much normative information follows from such an utterance. But with the normative sense of (2b), one gets the impression that philosophers who fail to care for the truth, or care for the truth in the right way, are normatively inadequate. Yet, while this flavor of normative generics is easy to identify, it is harder to characterize or explain. One way to get a purchase on it is by observing the discourse effects of utterances of utterances of normative generics. To this end, philosophers and linguists often discuss this in speech act-theoretic terms, pointing out that they have a hortative or prescriptive force. (Leslie 2015; Haslanger 2011; Hesni 2021) It's been said that normative generics give instructions, provide advice, or prescribe appropriate behavior. We concur that these are important and central uses of normative generics, but argue that the analysis should broaden its focus on some of the wider ways that normative generics figure in discourse.

Normative generics as derogatives and threats

Here are some examples of what we have in mind. First, utterances of normative generics often involve derogation or depreciation. Consider a very conservative family member uttering the following to a working mother, when both know the latter has no intention or means to stop working:

    (3) Women stay home to raise families.

As an instruction or bit of advice not to work, this utterance would be otiose. Felicity conditions for instructions tend to include expectations that the addressee has an interest in following the advice, or that there is a reasonable expectation that they will follow it. (3) would be a felicitous utterance even with a clear understanding that the addressee neither would nor could possibly follow the advice. A better way to understand this utterance of (3) in the envisioned context is as a means to derogate the addressee, pointing out the grounds for this derogation.

    Or, consider a coach addressing a young player sitting on the field crying after a painful tackle. The coach utters (1b). Again, though it is easy to detect the normative flavor of this utterance, and there is something hortative or prescriptive about it, we find unhelpful to think of such an utterance as, e.g., a command not to cry. (Commands tend to be infelicitous when it is common ground that the addressee won't comply with it. Cf. Mandelkern 2019.) Instead, we think it is more promising to understand the discourse function of such an utterance as a threat to the status of the addressee.

Scripts and definitions

We maintain that these kinds of uses of normative generics—as a means of derogation, or as a giving voice to threats—are fairly central to normative generics, yet the main accounts of normative generics (for example, those of Burton-Roberts 1976, 1977; Cohen 2001; Leslie 2015; Hesni 2021) wouldn't adequately explain why this is so.

    After making the case for these central uses of normative generics, the rest of our talk focuses on bringing together theoretical resources that allow us to explain them. We follow a recent proposal by Berio and Musholt (2022) which argues that a function normative

generics discourse involves the activation of schemata and scripts (in the sense of Schank and Abelson 1975, 2013) in specific social situations. Berio and Musholt argue, following Bicchieri and McNally (2018), that failures to meet the expectations engendered by schemata and scripts often trigger reactive attitudes and a dispositions to attempt to reestablish compliance with these expectations.

We argue that combining this proposal together with the metalinguistic account of generics from Krifka 2013 allows us to see normative generic discourse as a means of negotiating and policing normative expectations about the relevant kinds, and this, in turn allows us to analyze the relevant uses described above as derogations and/or threats.

References

Berio, L. and K. Musholt (2022). How language shapes our minds: On the relationship between generics, stereotypes and social norms. Mind and Language 38(4), 944–961.

Bicchieri, C. and P. McNally (2018). Shrieking sirens: Schemata, scripts, and social norms. Social Philosophy and Policy 35(1), 23–53.

Burton-Roberts, N. (1976). On the generic indefinite article. Language 52(2), 427–448.

Burton-Roberts, N. (1977). Generic sentences and analyticity. Studies in Language. 1(2), 155–196.

Cohen, A. (2001). On the generic use of indefinite singulars. Journal of semantics 18(3), 183–209.

Haslanger, S. (2011). Ideology, generics, and common ground. In C. Witt (Ed.), Feminist Metaphysics, pp. 179–207. Springe.

Hesni, S. (2021). Normative generics: Against semantic polysemy. Thought. 10(3), 218–225.

Krifka, M. (2013). Definitional generics. In A. Mari, C. Beyssade, and F. D. Prete (Eds.), Genericity, pp. 372–389. OUP.

Leslie, S. (2015). "Hillary Clinton is the only man in the Obama administration": Dual character concepts, generics, and gender. Analytic Philosophy 56(2), 111–141.

Mandelkern, M. (2019). Practical moore sentences. Nous 55(1), 39–61. Schank, R. C. and R. P. Abelson (1975). Scripts, plans, and knowledge. In IJCAI, Volume 75, pp. 151–157.

Schank, R. C. and R. P. Abelson (2013). Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Hillsdale: Earlbaun Associates, Inc.

Attitudes de se and Communication

Olga Kozyreva
Independent researcher | olgakozyreva66@gmail.com

In this talk, I focus on the problem of *de se* reports, i.e., the sentences of the form (1) 'Olga expects to travel to Prague'. The interpretation of such sentences assumes that Olga not only has some attitude about herself but also is aware that this attitude is about herself, and she may express it using the sentence (2) 'I expect to travel to Prague'. Usually, the explanation of this interpretation involves the fact that (1) has a silent anaphoric element PRO indicating *de se* reading: 'Olga expects PRO to travel to Prague' (Chierchia, 1966; Higginbotham, 2009). Usually, *de se* reports contrast with *de re* reports, i.e., the sentences of the form (3) 'Olga expects that she will travel to Prague'. In this case, one cannot be sure that Olga will agree to use sentence (2) since she may not be aware of the identity between herself and the object (*res*) she has an attitude about. Unlike sentences of the form (1), sentence (3) exhibits *de re / de se* ambiguity because nothing in its structure can help us to determine whether Olga is aware of the identity. A number of well-known examples in the literature are intended to demonstrate how lack of awareness is possible and to argue for

the necessary distinction between *de re* and *de se* reports (Castañeda, 1966; Lewis, 1979; Perry, 1979).

Although the problem of *de se* reports is primarily a linguistic problem, it is connected with philosophy of mind too. The main reason for arguing for the interconnection between the two fields lies in a so-called 'Reflection Assumption', i.e., an assumption that specific features of *de se* reports reflect some facts about the nature of first-person thought. My aim here is to question the 'Reflection Assumption' and to propose an alternative explanation of the difference between *de se* and *de re* reports.

First, I briefly overview the discussion on the problem. Since my main focus is on *de se* reading of the control complements with the PRO element, I talk more about Higginbotham's reflexive account and its limitations, such as neglecting the difference between explicit and implicit *de se* thoughts (Recanati, 2007).

Second, I critically discuss an attempt to modify Higginbotham's account by connecting the experiencer theta-role with the concept of 'minimal self' (Fiorin & Delfitto, 2014). This attempt, in my view, is one of the most vivid examples of accounts implicitly based on the 'Reflection Assumption'. I find it problematic because Fiorin & Delfitto conceive the minimal self as a mental entity accessible to an individual only by introspection. Mapping the semantics of natural language to such private entities brings us back to Frege's difficulty with private thoughts (Frege, 1956) and jeopardizes the possibility of communication.

Finally, I sketch my account of *de se* reports based on the notion of communicative aim, i.e., a specific aim of communication each participant of a communicative situation has in mind. In that case, *de se* reports differ from *de re* reports because of different communicative aims speakers try to achieve in communication. If the speaker feels the need to communicate that an individual she is talking about is willing to use a sentence of the form (2), then the speaker will use a disambiguated sentence like (1). However, if this information is irrelevant to communication, the speaker may use sentences like (3). So, I try to demonstrate that *de se / de re* ambiguity has nothing to do with metaphysics of first-person thought and can be fully reduced to the questions of theory of communication.

References:

Castañeda, H.-N. (1966). ""He": A Study in the Logic of Self-consciousness". In: Ratio 7.2, pp. 130–157.

Chierchia, G. (1966). "Anaphora and Attitudes *de se*". In: R. Bartsch, J. Van Benthem, P. Van Emde Boas (eds.) Semantics and Contextual Expressions. Foris, Dordrecht 1989, pp. 1–31.

Fiorin, G., Delfitto, D. (2014). "Control, Attitudes *de se* and Immunity to Error Through Misidentification". In: Rivista internazionale di filosofia e psicologia 5.2, pp. 184–206.

Frege, G. (1956). " The Thought: A Logical Inquiry". In: Mind 65. 259, pp. 289–311.

Higginbotham, J. (2009). "Remembering, Imagining, and the First Person". In: J. Higginbotham Tense, Aspect, and Indexicality. Oxford: Oxford University Press, pp. 212–245.

Lewis D. (1979). "Attitudes *De Dicto* and *De Se*". In: The Philosophical Review 88. 4, pp. 513–543.

Perry, J. (1979). "The Problem of Essential Indexicals". In: Noûs 13.1, pp. 3–21.

Recanati, F. (2007). Perspectival Thought. A Plea for Moderate Relativism. Oxford: Oxford University Press.

Tuesday, 27/Aug/2024 11:30 – 12:10

Philosophy of Language

Morphemes on Words

Kamil Lemanek

Maria Curie-Skłodowska University, Poland | kamil.lemanek@mail.umcs.pl

Kaplan's seminal "Words" (1990) in many ways initiated a new kind of consideration of the nature of words in philosophy, helping to establish an area of research that has garnered considerable interest in recent years. The ontology of words in particular has seen new positions and discussions arising, from novel artifactualist treatments (Irmak 2019), to structuralist accounts (Neftd 2019), to bundle-based approaches (Miller 2021). One exchange that has since also become influential is Hawthorne and Lepore's "On words" (2011), which addressed Kaplan's original work and leveraged morphology in arguing against the common-currency position he defended. It features their well-known "anti-anti-missile" example, one that invoked productive morphology to demonstrate a set of transparent and yet uninstantiated words. Though interesting and in many ways compelling, their account struggles with producing a strong and tenable distinction between morphemes and words, which their approach explicitly depends upon.

The aim of this talk is elaborate on the shortcomings in Hawthorne and Lepore's treatment of that distinction before turning to adapting the criteria they set out in order to improve it, introducing well-known considerations from the word-formation literature proper to linguistics (see, e.g., Matthews 1991; Murphy 2010; Plag 2005). This is done by first outlining the difference between free and bound morphemes, which illustrates that the original proposal of what seems to be a basic distinction between words and morphemes is untenable, as there is clear overlap between morphemes and words in the form of, e.g., free morphemes like "cat" or "man." That observation is then used to motivate an alternative distinction, namely, one between combinations of free and bound morphemes that do not render words and combinations of free and bound morphemes that do.

Within the scope of this new more workable distinction, Hawthorne and Lepore's original criteria relating to movability and uninterruptability are combined and joined into a single hybrid criterion, such that if a morpheme or combination of morphemes cannot be moved in the context of a reformulation of any simple sentence *salva congruitate*, then it cannot be a word. Where Hawthorne and Lepore initially demonstrated their own position with simple test cases, the same can be done with the hybrid criterion, such that it passes the original cases that Hawthorne and Lepore advanced without issue – e.g., relocating "men" in "do men like women?" or "unhappy" contrasted with "~happyun" (Hawthorne and Lepore 2011, 456). The new criterion and distinction integrate with their broader account, albeit it replacing the original, more basic distinction.

This is a somewhat unusual but interesting interdisciplinary task at the intersection of philosophy of language, linguistics, and ontology, and one that is important to the broader discussion on the ontology of words, not just in terms of improving upon an existing position, but in terms of providing a clearer picture of how morphemes and words may relate to one another. It is a small but very deliberate step in the direction of greater interdisciplinary engagement between philosophy of language and linguistics broadly construed, and a focused example of the need for that kind of work.

Reference List

Hawthorne, J., and Lepore, E. (2011). On words. *The Journal of Philosophy*, 108 (9), 447-485.

Irmak, N. (2019). An ontology of words. *Erkenntnis*, 84, 1139-1158.

Kaplan, D. (1990). Words. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 64, 93-119.

Matthews, P. (1991). Morphology: An introduction to the theory of word structure. London: Cambridge University Press.

Miller, J. T. M. (2021). A bundle theory of words. *Synthese*, 198, 5731-5748.

Murphy, L. M. (2010). *Lexical meaning*. Cambridge: Cambridge University Press.

Nefdt, R. (2019). The ontology of words: A structural approach. *Inquiry*, 62 (8), 877-911.

Plag, I. (2003). *Word-formation in English*. Cambridge: Cambridge University Press.

Monday, 26/Aug/2024 17:50 – 18:30
Philosophy of Mind
Between Agency and Passivity: Unpacking the Agential Status of Mind Wandering
Yizhi Li

Ruhr Universität Bochum, Germany | yizhi.li24@gmail.com

While it is commonly accepted that mind wandering possesses certain passive and non-agential dimensions, empirical research indicates that some facets of mind wandering may indeed be expressions of human mental agency (Klinger, 2009; Stawarczyk, 2018; See also Irving, 2021). Despite this intriguing possibility, philosophical inquiry into the agential status of mind wandering remains surprisingly sparse. Critical questions remain largely unaddressed: does mind wandering manifest our agency, and if so, in what manner and to what extent? This paper seeks to explore these questions, drawing on the latest scientific research on mind wandering and theoretical resources from philosophy of mental action (Brent & Titus, 2022; O'Brien & Soteriou, 2009) to examine the complex relationship between mind wandering and human agency. Specifically, it will critically evaluate various theoretical positions on mind wandering's agential status using empirical findings and philosophical analysis to identify and argue for a nuanced perspective that addresses existing gaps in the literature.

In the philosophical literature, four potential positions regarding the agential status of mind wandering, ranging from viewing mind wandering as fully agential to considering it entirely non-agential, can be identified:

1. Very strong Agentialism (O'Shaughnessy, 2000): Mind wandering is fully agential. According to O'Shaughnessy, wakeful consciousness (in contrast to dreaming and other global states of consciousness) is essentially agential. Therefore, mind wandering must be agential. "The mind remains intentionally active under headings at each point in that process [mind wandering]" (O'Shaughnessy, 2000, p. 217).

2. Strong Agentialism: Mind wandering is agential when certain types of mental activities are involved. One standard approach to the problem of the scope of mental action is to determine the agential status of one specific type of mental activity by analyzing its intrinsic structure. Philosophers who adopt this approach to the scope of mental action (e.g., Arango-Muñoz & Bermúdez, 2018; Peacocke, 2007; Soteriou, 2013) potentially support strong agentialism regarding mind wandering.

The general structure of the argument:

P1 A specific type of mental activity (such as imagining, reasoning, remembering) is constitutively agential.

P2 That type of mental activity can occur during mind wandering.

C Therefore, mind wandering is agential when it involves that type of mental activity.

3. Weak Agentialism (Irving, 2021): Mind wandering exhibits agential characteristics that are commonly associated with agency, suggesting that it embodies, at the very least, minimal forms of agency.

"Causally motivated (personal goal-relevant) mind-wandering shows us that minimal forms of agency can exist without guidance" (Irving, 2021, p. 626).

4. Anti-Agentialism (Metzinger, 2013, 2015, 2018): Mind wandering is a non-agential phenomenon.

The basic argument:

P1 Mind wandering, like breath and heartbeat, is a sub-personal phenomenon.

P2 Sub-personal phenomena cannot be the manifestation of human agency.

C Mind wandering is a non-agential phenomenon.

After mapping out the landscape of existing perspectives on mind wandering, I will first critically review and argue against all of the positions mentioned above and then introduce new arguments in support of a nuanced position labeled 'Very Weak Agentialism', which finds its place between Anti-Agentialism and Weak Agentialism, and claims that not all cases of causally motivated mind wandering reflect human agency, but certain cases indeed embody it. Central to the Very Weak Agentialism is the idea that the agential status of mind wandering is inherently context-dependent, a thesis aligns with the context regulation hypothesis in empirical studies on mind wandering (Andrews-Hanna et al., 2014; Smallwood & Schooler, 2015, p. 498). This hypothesis posits that the outcomes of mind wandering, be they positive or negative, are significantly shaped by the situational context. I will demonstrate how this context dependency not only supports the classification of Very Weak Agentialism but also enriches our understanding of the concept of mental agency.

In the concluding part, I aim to merge the insights obtained from the philosophical analysis of the agential status of mind wandering with the ongoing discussions about the psychological mechanisms underpinning mind wandering, particularly focusing on the debate surrounding cognitive control's role in mind wandering (McVay & Kane, 2010; Smallwood & Schooler, 2006). It will be highlighted how a refined philosophical perspective on the agency involved in mind wandering enriches our psychological grasp of the mechanisms underlying mind wandering. My conclusion will be that the agential part of mind wandering requires the involvement of cognitive control.

References

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, *1316*, 29–52. https://doi.org/10.1111/nyas.12360

Arango-Muñoz, S., & Bermúdez, J. P. (2018). Remembering as a Mental Action *. In *New Directions in the Philosophy of Memory*. Routledge.

Brent, M., & Titus, L. M. (Eds.). (2022). *Mental Action and the Conscious Mind*. Routledge. https://doi.org/10.4324/9780429022579

Irving, Z. C. (2021). Drifting and Directed Minds: The Significance of Mind-Wandering for Mental Agency. *The Journal of Philosophy*, *118*(11), 614–644. https://doi.org/10.5840/jphil20211181141

Klinger, E. (2009). Daydreaming and fantasizing: Thought flow and motivation. In *Handbook of imagination and mental simulation* (pp. 225–239). Psychology Press.

McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin*, *136*(2), 188–197. https://doi.org/10.1037/a0018298

Metzinger, T. (2013). The myth of cognitive agency: Subpersonal thinking as a cyclically recurring loss of mental autonomy. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00931

Metzinger, T. (2015). M-Autonomy. *Journal of Consciousness Studies*, *22*, 270–302.

Metzinger, T. (2018). Why Is Mind-Wandering Interesting for Philosophers? In K. C. R. Fox & K. Christoff (Eds.), *The Oxford Handbook of Spontaneous Thought*. Oxford University Press.

O'Brien, L., & Soteriou, M. (Eds.). (2009). *Mental Actions*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199225989.001.0001

O'Shaughnessy, B. (2000). *Consciousness and the World*. Oxford University Press. https://doi.org/10.1093/0199256721.001.0001

Peacocke, C. (2007). Mental action and self-awareness (I). In B. P. McLaughlin & J. Cohen (Eds.), *Contemporary Debates in Philosophy of Mind*. John Wiley & Sons.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946–958. https://doi.org/10.1037/0033-2909.132.6.946

Smallwood, J., & Schooler, J. W. (2015). The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology*, *66*(1), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Soteriou, M. (2013). *The Mind's Construction: The Ontology of Mind and Mental Action*. Oxford University Press.

Stawarczyk, D. (2018). Phenomenological properties of mind-wandering and daydreaming: A historical overview and functional. *The Oxford Handbook of Spontaneous Thought: Mind-Wandering, Creativity, and Dreaming*, 193.

Tuesday, 27/Aug/2024 17:30 – 18:10
Philosophy of Mind

On Propositional Attitudes as Natural Kinds
Agata Machcewicz-Grad

University of Warsaw, Poland | a.machcewicz-grad@uw.edu.pl

The question of whether propositional attitudes (PA), especially beliefs, can be considered natural kinds becomes more pressing in the context of recent popularity of superficialist accounts of PA. Authors from the superficialist camp argue that concepts of PA could never prove to 'carve nature at its joints'. That is, they could not be causally efficacious, concrete mental states held by subjects independently from observer. Instead, they are claimed to be context-dependent states attributed by interpreters depending on their pragmatic interests (Currie 2021; Schwitzgebel 2021; Posłajko 2024).

Before accepting the claim that PA cannot form natural kinds, one should consider more closely what conditions PA should fulfil in order to be regarded as legitimate candidates for natural kinds. This issue has not been often discussed explicitly (but see Perez 2004; Posłajko 2024). This paper is supposed to provide a more exhaustive analysis of possible readings of the view that PA form natural kinds.

I argue that, contrary to what superficialists claim, folk psychological ascriptions of PA carry substantial ontological commitments. The implicit assumption behind folk psychological practices is that the attributions of PA, if done accurately, provide explanations of the target's observable behaviour in terms of concrete, unobservable mental causes. They assume, moreover, that these attributions are made on the basis of ceteris paribus generalisations that describe the nature of the mental states being ascribed. Since the theoretical purpose of psychology is to reveal the laws governing human mental states, the folk psychological categories can be considered as referring to the same mental states to which psychologists refer. Both lay people and psychologists talk about the same states even if they associate different descriptions with the terms they use. The former, presumably, defer to the latter the task of revealing the underlying nature of psychological

kinds and, possibly, correcting the folk psychological generalisations. If such an account proves plausible, then they pose a threat for superficialist accounts of PA.

The proposed account of PA shall be situated in the context of competing theories of natural kinds. In the past decades there was a vivid discussion about natural kinds in philosophy of science. It started from radical physicalist (Lewis 1983) and essentialist (Putnam 1975) theories to give rise, later on, to more pluralistic theories. Lewis's physicalist account was to the effect that only fundamental physical kinds can be considered natural kinds simpliciter. All the other kinds posited by special sciences can be more or less imperfectly natural depending on whether - and how smoothly - they can be reduced by means of definitions to the perfectly natural, physical kinds. Mental kinds are not good candidates for natural kinds when considered in light of Lewis' criteria, chiefly due to their multiple realizability. However, there are independent reasons not to consider Lewis' account plausible, having to do with questionable status of physicalist premises and non-reducibility of natural kinds posited by various special sciences, e.g. biology.

Natural kind essentialism in its traditional formulation does not have to be reductive, but it requires essential properties of a kind to be intrinsic properties. Since this requirement is problematic not only with regard to the posits of social sciences which define their terms by means of historical, relational and functional properties, but also to paradigmatic biological kinds, it has been claimed by many authors to be unduly restrictive. A possible strategy to preserve essentialism while admitting those latter kinds as natural kinds is to accept essentialistic definitions involving extrinsic, apart from intrinsic, properties (Okasha 2002). If this strategy proves plausible, then the prospects for accepting psychological natural kinds look significantly better.

More liberal, non-essentialist theories of natural kinds have also been proposed. Among the best developed ones is the 'simple causal account' by Khalidi (2013) according to which natural kinds are to be distinguished on the basis of projectibility, explanatory efficacy and predictive value (Khalidi 2013: 55). The account is pluralistic, since it rejects reductionist requirements and allows for non-intrinsic properties as distinctive properties of natural kinds. It also abandons the essentialistic requirement that the boundaries between kinds are to be clear-cut and discrete. The causal theory proposed by Khalidi does not give any substantial restrictions which would render mental states improper candidates for natural kinds. It admits, moreover, that some folk categories might prove to be referring to natural kinds.

It stems from the overview of the main theories of natural kinds that at least some of those theories do not exclude the possibility that PA form natural kinds. The more restrictive accounts like Lewis' physicalist account or the intrinsic-properties essentialism do exclude this possibility, but there are good independent reasons not to adopt these accounts. The more pluralistic theories like extrinsic-and-intrinsic-properties essentialism and simple causal theory admit the possibility of PA being natural kinds (although neither of them settles if they actually are).

Superficialists about propositional attitudes argue as if the question of whether PA can be natural kinds was settled definitely. It seems, however, that this question is far from being settled. The answer to this question differs depending on more general issues having to do with metaphysics of natural kinds. If PA as they are used in folk psychology may refer to the mental natural kinds explored by psychology, then superficialists are in trouble.

References

Currie, D. S. (2021), How beliefs are like colors. Synthese, 199 (3-4), 7889-7918.

Khalidi, M. A. (2013). Natural Categories and Human Kinds: Classification in the Natural and Social Sciences. New York: Cambridge University Press.

Lewis, D. K. (1983). New work for a theory of universals. Australasian Journal for Philosophy, 61 (4), 343-377.

Okasha, S. (2002), Darwinian Metaphysics: Species and the Question of Essentialism. Synthese. 131 (2), 191–213.

Pérez, D.I. (2004). Mental concepts as natural kind concepts. Canadian Journal of Philosophy, 30 (sup. 1), 201-225.

Posłajko, K. (2024). Unreal beliefs. Bloomsbury Publishing.

Putnam, H. (1975). The meaning of 'meaning'. Minnesota Studies in the Philosophy of Science, 7, 131-193.

Schwitzgebel, E. (2021). The Pragmatic Metaphysics of Belief. In. C. Borgoni, D. Kindermann, A. Onofri (Eds.), The Fragmented Mind. Oxford University Press.

Monday, 26/Aug/2024 16:30 – 17:10
Philosophy of Mind
Perceiving and Misperceiving Properties
Giulia Martina
University of Konstanz, Germany | giulia.martina@tutanota.com

What does it take to perceive a sensible property, such as an object's colour? It has been argued that there conditions on perceiving a property which are not conditions on perceiving an object (e.g. Millar 2022, 2023). We may count as seeing an object no matter how it looks to us or how little we can tell about it; but seeing its colour is, the view holds, more demanding. If a blue bead in pink light looks black, we surely see the bead, but do wesee its colour? If intuitions diverge, how do we decide?In this paper, I present a challenge to any conditions on perceiving an object's property that appeal either to how the object appears to us or to what we can judge about it. The challenge turns on the difficulty of drawing a line between cases of ordinary perceptual variation, where everyone agrees that we perceive the target property even though it looks a bit different across conditions, and cases where perception is deeply misleading as to the nature of the property, as with the black-looking bead. There are many kinds of failures involved in property perception which do not fit the traditional category of illusion and are not cases of failing to perceive the target property(e.g. Kalderon 2011, Macpherson & Batty 2016, Alford-Duguid 2020). Minimal conditions on property perception allow for a better account of both ordinary perceptual variation and the variety of failures in property perception.I conclude by outlining a distinction between perceiving as a perceptual relation andattributions of states of perceivinga property P to subjects (e.g. 'she could see the bead's colour')which we make in everyday discourse. This distinction, I suggest, helps us make sense of the diverging intuitions about limiting cases like the black-looking blue bead.

Wednesday, 28/Aug/2024 9:40 – 10:20
Philosophy of Language
The Fregean Hierarchy
Mark Philip McCullagh
University of Guelph, Canada | mmcculla@uoguelph.ca

Many philosophers continue to be attracted to a Fregean (1892) approach to indirect discourse and attitude ascription (e.g. recently Yalcin 2015; Lederman 2022). But there is a hanger-on accompanying that theory, the debate over which is at something of a standoff. It is the "hierarchy view": that for each word there is an infinite hierarchy of semantic values

that it contributes in different kinds of occurrence. Frege does not mention this idea when explaining his theory, but many readers have taken his theory to entail it. Saul Kripke (2008, 183) even called it a "familiar consequence" of the theory. Others deny the entailment, but nonetheless maintain that the hierarchy view is a natural extension of Frege's stated doctrine, to cover cases that he didn't originally have in mind. Those are cases of so-called "multiple embeddings," as in

Susan thinks that John believes that Paris is beautiful.

The idea is that since "Paris" is doubly embedded by indirect discourse verbs, its semantic value in that occurrence is not its customary sense—which would be its value if it were only singly embedded—but its "indirect" sense, a distinct sense that "presents" its customary sense. And so on, for higher degrees of embedding: when embedded to degree n, a word's semantic value is its level n sense.

Almost all of the discussion of this view has concerned its bare tenability rather than its plausibility. My concern here is with the latter. I will examine what seem to me to be the two strongest considerations given in support of the view.

One consideration appeals to cases of multiple embeddings in which intersubstitutability salva veritate seems to be more stringent than it is within single embeddings (Mates 1950). The claim is that the hierarchy explains this apparent phenomenon. For it allows that two words can share their level n senses while having different level n + 1 senses. In such a case, the words are guaranteed to be intersubstitutable when embedded to degree n, but not so guaranteed when embedded to degree n + 1. Many philosophers (e.g. Fine 2007; Horwich 2014) have endorsed this idea. But there is a methodological problem: this is not the only possible explanation of the Matesian phenomenon. There is another sort of explanation which involves no hierarchy. So the hierarchy view does not come close to being supported by an inference to the best explanation.

There are also arguments meant to show that the hierarchy view follows from Frege's semantic theory. The best argument of this sort is due to Tyler Burge (1979; 2005). Burge argues for the hierarchy on the basis of claims about how Frege's core principle plays out in the formalization of a natural language. His argument is careful and detailed—dauntingly so. It is widely endorsed (e.g. Salmon 2005). Its crucial step rests on a claim about the well-formedness of expressions in a "logically perfect" (Frege 1892, 58) language. On examination, it turns out that that claim is plausible only if the hierarchy view is true. The argument, then, begs the question, and gives the hierarchy view no support.

So the hierarchy view has even less going for it than has been thought. What then explains its enduring appeal—such as it is, mostly among Fregeans? How is it that so many readers, even while differing in their alliegance to Frege's overall approach, have thought that his view at least strongly suggests the hierarchy? That common reaction is not due to some complex argument such as Burge's. What then is behind it?

I diagnose it as resting on an error in thinking of the base case of Fregean reference shift. The error is mistaking what a word's semantic value does, for something the word itself does. When embedded, an attitude or discourse verb does not *induce* embedding, precisely because its semantic value in such an occurrence is not what it is when it occurs unembedded (and does induce embedding). On the Fregean approach, understood correctly, there is *no such thing* as "multiple embedding"—let alone a theory of it. This illustrates again the power of Frege's dictum "never to ask for the meaning of a word in isolation, but only in the context of a sentence" (1884, xxii).

References

Burge, Tyler. 1979. Frege and the hierarchy. *Synthese* 40:265–81.

———. 2005. Postscript to "Frege and the hierarchy". In *Truth, thought, reason: essays on Frege*, 167–210. Oxford: Clarendon Press.

Fine, Kit. 2007. *Semantic relationism*. Oxford: Blackwell.

Frege, Gottlob. 1884. *The foundations of arithmetic.* Translated by J. L. Austin. Evanston, IL: Northwestern University Press.

———.1892. Über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*100: 25–50. Translated as "On sense and meaning" in *Translations from the philosophical writings of Gottlob Frege*, 2d edition (1960), ed. and trans. Peter Geach and Max Black, 56–78. Oxford: Blackwell.

Horwich, Paul. 2014. Critical notice of *Seven Puzzles of Thought and How to Solve Them: an Originalist Theory of Concepts*, by R. M. Sainsbury and Michael Tye. *Mind* 123:1123–39.

Kripke, Saul. 2008. Frege's theory of sense and reference: Some exegetical notes. *Theoria* 74: 181–218.

Lederman, Harvey. 2022. Fregeanism, sententialism, and scope. *Linguistics and Philosophy* 45: 1235–75.

Mates, Benson. 1950. Synonymity. *University of California Publications in Philosophy* 25:201– 26.

Salmon, Nathan. 2005. On designating. *Mind* 114(456):1069–1133.

Yalcin, Seth. 2015. Quantifying in from a Fregean perspective. *Philosophical Review* 124:207–53.

Tuesday, 27/Aug/2024 11:30 – 12:10

Philosophy of Mind

A Challenge for the Acquaintance-Based Conception of Inner Awareness

Jakub Mihalik

Czech Academy of Sciences, Institute of Philosophy, Czech Republic | mihalik@flu.cas.cz

Up until recently, the prevalent construal of 'inner awareness', a subject's special awareness of their own mental states which renders these states conscious, was representationalist. According to Rosenthal (2005), for example, one's mental state is conscious due to one's having a suitable higher-order thought (HOT) that represents it; according to Kriegel's (2009) self-representationalist account, on the other hand, mental states are conscious due to representing themselves. If, however, inner awareness of our mental states is a matter of representing these states, it would seem that we are kept 'cognitively apart' from our own conscious mental contents, which seems implausible. Since the representational states that constitute inner awareness – and thus determine the what-it's-likeness of experience, according to the representationalists – could misrepresent the mental states that this awareness makes conscious, one could, for example, experience a green bus while being in a 'red-bus-ish' conscious perceptual state, or in no conscious perceptual state at all (see Coleman 2015).

This unwelcome 'cognitive distance' between ourselves and our conscious states that representationalism brings about (see, however, Kriegel 2009), has led some proponents of inner awareness (e.g. Williford 2015, Duncan 2018, Levine 2019; Giustina (forthcoming)) to suggest that inner awareness is best understood in terms of Russell's notion of acquaintance. For Russell (1910–11), acquaintance is a direct, non-conceptual cognitive relation between a subject and an intentional object, where the object is (immediately) presented to the subject. An acquaintance-based model of inner awareness would arguably cancel any cognitive distance between a subject and her conscious states as these would be directly presented to her, instead of represented, overcoming the described misrepresentation problem. Such a model, however, may seem to face a challenge concerning the naturalisability of consciousness since acquaintance – unlike representation – is often viewed as naturalistically suspect, or even mysterious.

While some proponents of acquaintance have bitten the bullet and endorsed a non-naturalistic conception thereof (Levine 2019), others have argued that acquaintance is compatible with naturalism. I shall investigate arguably the best developed naturalistic model of acquaintance, suggested by Coleman (2019), who models acquaintance in terms of a special, non-representationalist version of the HOT theory. For Coleman, then, our inner awareness of our mental states is a matter of having HOTs about these states, where these HOTs do not represent our mental states, but rather 'quote' them, due to the target mental states' being embedded within the HOTs (see Coleman 2015, 2018). Coleman's quotational-HOT (or simply QHOT) model, if viable, would thus avoid the misrepresentation problem and would accommodate our sense that in being phenomenally conscious, we are in direct, substantial and near infallible cognitive contact with our mental (for example, sensory) states, that the notion of acquaintance aims to capture.

After exploring the key features of Coleman's conception, I shall attempt to raise a challenge for it that is inspired by Levine's (2007) critique of Balog's related 'quotational account' of phenomenal concepts. Balog's approach, according to Levine, fails to explain the specific 'cognitive presence' of phenomenal states which our direct phenomenal concepts (those concepts we acquire by directly attending to what it is like to be in various phenomenal states, Chalmers 2003) seem to afford us. According to Levine (2007, p. 159), a phenomenal state is cognitively present for an organism just in case the organism enjoys an especially immediate and intimate cognitive access to this state. I'll argue that a related objection applies to Coleman's model of acquaintance, despite significant differences between his and Balog's accounts.

I'll first explain that acquaintance with a mental state makes it cognitively present for a subject in Levine's special sense. Then I'll argue that Coleman's QHOT account, despite plausibly avoiding the pitfalls of representationalism, falls short of explaining this cognitive presence. While our QHOTs plausibly cognitively relate us to the target mental states in being intentionally 'about' these states, I shall explain that this intentional relation cannot, on its own, fully account for cognitive presence, in Levine's intimacy-involving sense, since standard representing HOTs are intentionally 'about' their target states too, without rendering these states cognitively present in this sense.

At this point, the proponents of the QHOT account are likely to appeal to the physical presence of the target mental state within the QHOT, but I shall argue that it's unclear why this physical presence should result in the state's cognitive presence for the organism that acquaintance seems to involve. Mere physical presence, after all, seems compatible with the mental state's not having any physical impact on the QHOT frame, which strongly suggests that it has no cognitive impact, that would go beyond an intentional relation, on the organism either. But, as explained above, merely being intentionally related to the target state seems insufficient for cognitive presence of the state for the organism, hence for acquaintance. If my critique is plausible, the QHOT account faces fatal problems, which should motivate us to search for theoretical alternatives.

References:

Balog, K. (2012). Acquaintance and the mind-body problem. In S. Gozzano & C. S. Hill (Eds.), New Perspectives on type identity: The mental and the physical. Cambridge: Cambridge University Press.

Coleman, S. (2015). Quotational higher-order thought theory. Philosophical Studies, 172(10), 2705–2733.

Coleman, S. (2018). The merits of high-order thought theories. Trans/Form/Ação 41, pp. 31– 48.

Coleman, S. (2019). Natural acquaintance. In J. Knowles & T. Raleigh (Eds.), Acquaintance: New essays (pp. 49–74). Oxford: Oxford University Press.

Duncan, M. (2018). Subjectivity as self-acquaintance. Journal of Consciousness Studies, 25 (3– 4), pp. 88–111.

Giustina, A. (forthcoming). Inner Acquaintance Theories of Consciousness. Oxford Studies in Philosophy of Mind 4.

Kriegel, U. (2009). Subjective consciousness: A self-representational theory. Oxford: Oxford University Press.

Levine, J. (2007). Phenomenal concepts and the materialist constraint. In T. Alter & S. Walter (Eds.), Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism (pp. 145–166). New York: Oxford University Press.

Levine, J. (2019). Acquaintance is consciousness and consciousness is acquaintance. In J. Knowles & T. Raleigh (Eds.), Acquaintance: New Essays (pp. 33–48). Oxford: Oxford University Press.

Rosenthal, D. M. (2005). Consciousness and mind. Oxford: Oxford University Press.

Russell, B. (1910–11) Knowledge by acquaintance and knowledge by description. Proceedings of the Aristotelian Society, New Series, Vol. 11 (1910 - 1911), pp. 108-128

Williford, K. (2015). Representationalisms, subjective character, and self-acquaintance. In T. Metzinger & J. M. Windt (Eds). Open MIND: 39(T). Frankfurt am Main: MIND Group. doi: 10.15502/9783958570054

Monday, 26/Aug/2024 17:10 – 17:50
Philosophy of Language
The Frege-Strawson Approach to Presuppositions Reconsidered
Andrei Moldovan

University of Salamanca, Spain | mandreius@usal.es

It is common to find in the literature on presuppositions a discussion of the so-called "Frege-Strawson" approach, either specifically in relation to singular terms such as definite descriptions or in general with respect to the phenomenon of presupposition as such (e.g., Kaplan 1970: 279; Kripke 1977: 269; Garcia-Carpintero 2000: 13; Salmon 2007: 69; Pelletier & Linsky 2005: 203; Elbourne 2013: 2; Schoubye 2013; Beaver, Geurts & Denlinger 2021: §2). For instance, Elbourne (2013: 2) writes that "the Frege–Strawson theory" is the theory according to which "definite descriptions are basically referential and introduce a presupposition to the effect that there is a unique entity that satisfies the nominal descriptive content" (2013: 2). I wish to argue that the differences between the Fregean and the Strawsonian accounts are sufficiently important so as to motivate a separate treatment of them. Scholars have pointed out that the two approaches are different, and that both of them are somewhat underdetermined with respect to distinctions that are now commonplace in natural language semantics and pragmatics (see, more recently, the discussion in Sander 2021). But the main difference seems to me to be the following. What is standardly called the "Frege-Strawson theory" places presuppositions at the level of what Kaplan (1989) calls *character*. This is somewhat fare to Strawson's own approach, if we interpret it as a semantic proposal. Strawson's notion of *conventional meaning* is close to Kaplan's (1989) notion of *character*. According to Strawson (1950: 327), to talk about the meaning of an expression or sentence is to talk about the *conventions* governing its correct use. And Kaplan (1989: 520-524) characterizes character as a property of only to words and phrases as types, to which it is directly associated by convention. However, Frege's notion of *sense* is closer to Kaplan's (1989) notion of *content*, rather than to that of character. Kaplan observes that (1989: 502) his notion of *content* in inspired by Carnap's (1947) notion of *intension*, and this, in turn, is inspired by Frege's notion of *sense*. A presuppositional expression such as a proper name is for Frege simply an expression that has a sense, but the sense does not always determine a reference. However, as several authors have pointed out, Frege does not have a notion corresponding to the Kaplanian *character*, or

conventional meaning (Burge 1990: 244; Heck 2002: 7; Perry 1977: 9). That is why I find it inadequate to talk about the "Frege-Strawson theory of presuppositions".

In the second part of the paper, I consider the question whether the difference between the two approaches, the Fregean and the Strawsonian, has some bearing on the way in which presuppositions are represented in truth-conditional semantics. I look into the possibility of capturing this difference in the standard formal framework for truth-conditional semantics developed in Heim and Kratzer (1998), for extensional contexts, and Fintel and Heim (2011), for intensional contexts. The latter framework allows to draw the Kaplanian distinction between character and content. I take the presuppositions of definite descriptions as a case study. In Heim and Kratzer (1998: 80) formulate presuppositions as *partial functions*. On my interpretation, a Fregean account of presuppositions treats the intension of an expression which carries a presupposition as a partial function. On a Strawsonian account, the *character* of an expression which carries a presupposition is a partial function from contexts to intensions. I compare the predictions that the two theories for the analysis of sentences that contain DDs embedded in intensional contexts. I focus on the analysis of the *de dicto* readings of non-doxastic propositional attitudes ascriptions.

References

Beaver, D. I. and Geurts, B., Denlinger (2021) "Presupposition", *The Stanford Encyclopedia of Philosophy* , Edward N. Zalta (ed.).

Burge, Tyler (1990) "Frege on Sense and Linguistic Meaning", in D. Bell and N. Cooper (eds.), The Analytic Tradition (Oxford: Blackwell), 30–60.

Carnap, R. (1947) *Meaning and Necessity: a Study in Semantics and Modal Logic*. University of Chicago Press.

Elbourne, P. (2005) *Situations and Individuals*, Cambridge, MA: MIT Press.

Elbourne, P. (2013) *Definite Descriptions*, Oxford: Oxford University Press.

Frege, G. (1892) "Uber Sinn und Bedeutung", *Zeitschrift fur Philosophie und philosophische Kritik*, C, 25–50. English Translation: "On Sense and Meaning", in B. McGuinness (ed.), *Frege: collected works*, Oxford: Blackwell, 157–177.

Fintel, K. von and Heim, I. (2011) *Intensional Semantics*, retrieved from http://web.mit.edu/fintel/fintel-heim-intensional.pdf.

García-Carpintero, M. (2000) "A presuppositional account of reference fixing". *Journal of Philosophy* 97 (3):109-147.

Heck, Richard (2002). "Do demonstratives have senses?" *Philosophers' Imprint* 2 (2): 1-33.

Heim, I. and Kratzer, A. (1998) *Semantics in Generative Grammar*, Blackwell Textbooks in Linguistics, Oxford: Blackwell.

Kaplan, D. (1970) "What is Russell's Theory of Descriptions?", in W. Yourgrau and A. D. Breck (eds.), *Physics, Logic, and History*, Routledge, 277-295.

Kaplan, D. (1989) "Demonstratives – An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals", in J. Almog, J. Perry and H. K. Wettstein (eds.): *Themes from Kaplan*, Oxford: Oxford University Press, 524–526.

Kripke, S. (1977) "Speaker's Reference and Semantic Reference," *Midwest Studies in Philosophy*, 2, 255–276.

Pelletier, F.J. and Linsky, B. (2005) "What is Frege's Theory of Descriptions?" in B. Linsky & G. Imaguire (eds.) *On Denoting: 1905-2005*, Philosophia Verlag, 195-250.

Perry, John (1977) "Frege on Demonstratives", *The Philosophical Review* 86, no. 4: 474-97.

Sander, T. (2021). Understanding Frege's notion of presupposition. *Synthese*, 199 (5-6), 12603-12624.

Salmon, N. (2007). "About Aboutness". *European Journal of Analytic Philosophy* 3 (2):59-76.

Schoubye, A. (2013) "Ghosts, Murderers, and the Semantics of Descriptions", *Noûs* 47(3), 496-533.

Strawson, P. F. (1950) "On Referring", *Mind*, 59, 320–344.

Probabilistic Arbitrary Reference

Matteo Nizzardo

University of St Andrews | mn85@st-andrews.ac.uk

Arbitrary Reference is the idea that we can refer to individual entities with some degree of arbitrariness. Although there are different accounts of Arbitrary Reference currently on the market, nearly all of them can be challenged on the basis that they entail the existence of free-floating semantic facts, namely: semantic facts which are not grounded in any non-semantic fact.

To see this, let $w1$ and $w2$ be two possible worlds which agree with respect to all non-semantic facts. Suppose that $w1$ is such that Charlie, a first year student at Oxford University, writes down a proof to the effect that every natural number is either even or odd, starting it with the supposition "Let n be a natural number". By stipulation, $w2$ is also such that Charlie, a first year student at Oxford University, writes down a proof to the effect that every natural number is either even or odd, starting it with the supposition "Let n be a natural number". (Whether the Charlie in $w2$ is the same individual as the Charlie in $w1$, or just one of his counterparts, is irrelevant here. For this reason, I will here remain silent on the problem of transworld individuals and counterparts.) Since $w1$ and $w2$ agree with respect to all non-semantic facts, the context and the use facts associated with Charlie's proof in $w1$ must be the same as the context and the use facts associated with Charlie's proof in $w2$. The challenge arises because Arbitrary Reference allows for the fact that the referent of the term 'n' in $w1$ might not be the same as the referent of the term 'n' in $w2$ --- despite all non-semantic facts are exactly the same in $w1$ and $w2$.

However, it is a common assumption in the Philosophy of Language that every semantic fact is grounded in some non-semantic facts. We find this assumption in virtually all theories of reference, be them internalist or externalist.

In this paper I propose a solution. First I argue that the friends of Arbitrary Reference can answer the challenge by appealing to the notion of *indeterministic grounding*. Indeterministic grounding happens when some low-level facts [P1] ... [Pn], which provide a full grounding base for some incompatible high-level facts [Q1] and [Q2], underdetermine which of [Q1] or [Q2] obtains. In this case, although [P1] ... [Pn] fully ground [Q1] and [Q2], the relation of grounding between low- and high-level facts is indeterministic: the obtaining of [P1] ... [Pn] alone doesn't suffice for the obtaining of [Q1] or [Q2]. Some form of chance is required, which however doesn't threaten the grounding chain: whichever of [Q1] and [Q2] obtains, it will be grounded, albeit indeterministically, on [P1] ... [Pn].

Then I propose a new account of Arbitrary Reference as a probabilistic phenomenon. Informally, the idea is that when we make suppositions like "Let n be a natural number'' we introduce some probabilistic constraints on the possible referents of the instantial term 'n'. I argue that this new account should be preferred over the classical versions of Arbitrary Reference for its ability to build a bridge between cases of canonical and arbitrary reference and the new insights it offers on the phenomenon of semantic vagueness.

Keeping Count of Co-Predications

Gareth Peter Gordon Norman

Massachusetts Institute of Technology, United States of America | garethn@mit.edu

Co-predications are a class of sentences thought to be a serious puzzle for truth-conditional semantics. Suppose that I use (1):

(1) The book, which was checked out of the library, was written by Kripke.

Whilst talking to you about my copy of Naming and Necessity. A standard truth-conditional semantics will say that the core of (1)'s meaning is:

(1) is true iff there exists exactly one x such that x is a book, x was checked out of the library and x was written by Kripke.

This is oversimplified, but it isn't entirely wrong. The point is that we ascribe two predicates, checked out of the library and written by Kripke to the referent of the book. The problem: naively, Kripke didn't write my copy of Naming and Necessity. It was printed by a machine. Naively, I didn't check Naming and Necessity} out of the library. It's a piece of informational content with no spatiotemporal properties. Therefore, the right-hand side of the bi-Conditional doesn't hold, and (1) is either false or a category mistake, contrary to intuition. In a series of recent papers and a book manuscript, Liebesman & Magidor (hereafter "L&M") address this puzzle by rejecting the naive metaphysics, instead defending:

Property Versatility View (PVV):

Co-predications are possible because properties aren't sortally restricted. (This is oversimplified. L&M merely think the sortal restrictions are much more relaxed than the standard view (L&M 2017, p149)

For example, book-types can be checked out of libraries if their tokens are. Resultantly, there's nothing incoherent about (1)'s truth-conditions.

PVV elegantly resolves the puzzle of co-predication. This only increases the burden to scrutinise it. That's the aim of this paper. I'm going to produce novel counting ambiguities (novel, at least, for the literature on co-predication), and use it to undermine PVV.

One of PVV's purported advantages is its handling of ambiguity in counting sentences. Consider:

One-Collection:

The only book written by Jane Austen checked out of the library is a copy of her collected works, call it *b*. That is, *b* is *Sense and Sensibility*, *Pride and Prejudice*, *Mansfield Park*, *Emma*, *Northanger Abbey*, *Persuasion* and *Lady Susan* all bound into a single volume (adapted from Asher 2008, p161)

How many books were checked out of the library in One-Collection? Intuitively there are two possible answers:

(2) One book written by Jane Austen was checked out of the library.

(3) Seven books written by Jane Austen were checked out of the library.

PVV predicts (2) and (3) have both true and false readings. Due to property-versatility, we get predicate extensions like:

[[book]] ={*b*, *Sense & Sensibility*, *Naming & Necessity*,...}
[[checked out of the library]] ={*b*, *Sense & Sensibility*, *Pride & Prejudice*,...}

Then in a context, we use *book* restrictedly, so we either count only book-tokens or only book-types (L&M 2017, p136)

In this paper, I consider another counting ambiguity and argue that it cannot be accommodated within PVV. The ambiguity in question are textit{event-related} readings of counting sentences. Consider:

Toll-Booth:

You work at a toll booth on a lock. One day, the ships Vesroth and Vargas I pass through the lock together. They both pay the toll. Later that day, Vargas I comes back through again, and she pays a second toll.

Krifka (1990) observed you can use sentences like:

(4) Three ships passed through the lock.

To report how many times a ship passed through the lock (hence the term "event-related"). We've counted Vargas I as two ships, despite being a single entity.

There are two extant approaches to event-related readings, and I consider the consequences of adding either to PVV. One one approach, *the Determiner View*, there is a silent determiner which measures event-participants (e.g. Krifka (1990), Doetjes & Honcoop (1997)). On another approach, *the Stage View*, we count stages (time-slices of individuals) instead of individuals (e.g. Barker (1990), Gotham (2021)). I consider:

Library:

The library has one copy of Jane Austen's collected works, b. Last week, b was checked out, returned, and then checked out again. No other books by Jane Austen were checked out.

Intuitively, the following:

(5) Fourteen books written by Jane Austen were checked out of the library.

Is not true in Library. I argue that on both the Determiner View and Stage View, PVV incorrectly predicts that (5) has a true reading. PVV tells us that book-types are books, written by Jane Austen and can be checked out of libraries. So on the Determiner View, we get the reading because there are two distinct events where 7 books written by Austen are checked out of the library (*Sense and Sensibility*, *Mansfield Park*, and so on). This yields a count of 14 books. On the Stage View, we get the reading because if *Mansfield Park* is checked out of the library twice, then there are at least two stages: *MP*-during-first-checkout, and *MP*-during-second-checkout. Extending this reasoning to Austen's other novels yields a count of 14 books.

After considering and rejecting some responses that could be made on behalf of PVV, I propose that we strive for a unified theory for counting sentences upon which token, type and event-related readings are derived. The upshot: counting ambiguities are largely an independent project. They should be part of the semantic and pragmatic toolkit for any theory of co-predication. PVV thus doesn't have the advantage with counting ambiguities it might be thought to have.

References

Asher, N. (2008). A Type Driven Theory of Predication with Complex Types. Fundamenta Informaticae, 84(2), 151–183. Retrieved June 22, 2023, from https://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=33410207&site=ehost-live&scope=site&authtype=shib&custid=s8978330

Barker, C. (1999). Individuation and Quantification. Linguistic Inquiry, 30(4), 683–691. JSTOR: 4179087. Retrieved November 10, 2023, from https://www.jstor.org/stable/4179087

Doetjes, J., & Honcoop, M. (1997). The Semantics of Event-Related Readings: A Case for Pair-Quantification. In A. Szabolcsi (Ed.), Ways of Scope Taking (pp. 263–310). doi:10.1007/978-94-011-5814-5 8

Gotham, M. (2021). Event-related readings and degrees of difference. Proceedings of Sinn und Bedeutung, 25, 325–339. doi:10.18148/sub/2021.v25i0.940

Krifka, M. (1990). Four thousand ships passed through the lock: Object-induced measure functions on events.

Linguistics and Philosophy, 13(5), 487–520. doi:10.1007/BF00627291

Liebesman, D., & Magidor, O. (2017). Copredication and Property Inheritance. Philosophical Issues, 27, 131–166. JSTOR: 26611258. Retrieved June 14, 2023, from https://www.jstor.org/stable/26611258

Challenging Khoo's Argument Against Ambiguity in Dogwhistles

Jesper Arent Leo Olsson
University of Graz, Austria | jesper.olsson@uni-graz.at

The carefully crafted language used by politicians often involves expressions known as *dogwhistles*. Dogwhistles are expressions that convey two messages: one that is neutral and one that is controversial or norm-violating, and where the latter is designed to appeal to specific subsets of the speaker's audience while remaining unnoticed by others. In this paper, I argue against Khoo's (2017) argument that dogwhistles cannot be ambiguous expressions, an argument which many have used in order to motivate fully pragmatic theories of how dogwhistles work. The secondary aim is to develop a tangible ambiguity view of dogwhistles that accommodates recent empirical findings on dogwhistles' community specific meanings.

A paradigm example of a dogwhistle is "inner city." During an interview on "Bill Bennet's Morning in America," Congressman Paul Ryan addressed poverty and welfare reforms by saying:

We have got this tailspin of culture, in our inner cities in particular, of men not working and just generations of men not even thinking about working or learning the value and the culture of work, and so there is a real culture problem here that has to be dealt with. (Lowery, 2014, bold is added)

In the literature, "inner cities" is ascribed the general meaning "central section of a city," but is also understood as conveying something about the race and status of the residents of the urban locations that are addressed, namely that one is talking about *poor African American city neighborhoods*. Khoo (2017) argues that "inner city" cannot be ambiguous because the expression fails tests for ambiguity such as *the contradiction test*. From this, Khoo contends that dogwhistles in general do not work via ambiguity. However, it is well known that not all ambiguous expressions reliably pass ambiguity tests such as the one Khoo uses. Polysemes, which are ambiguous expressions with two related meanings, are known to fail tests for ambiguity (Viebahn 2018; Geeraerts 1993 and Gillon 2004). I show that dogwhistles with related meanings systematically fail the contradiction test (e.g., "inner city" and "cultural enrichment") whereas dogwhistles with unrelated meanings actually pass the test (e.g., "jogger" and "skittles".) What we end up with are highly variable predictions, and the conclusion I draw is that the test underlying Khoo's argument is unreliable, and that we ought to remain open to the possibility that dogwhistles may work via ambiguity.

A recent empirical study by Boholm & Sayeed (2023) shows that certain Swedish dogwhistles undergo community-specific semantic shifts. By analyzing Swedish online forums, Boholm and Sayeed show that on forums linked with controversial ideologies, users consistently and systematically use certain terms with a dogwhistle meaning whereas users on other forums almost exclusively use the same terms with their neutral meaning. To accommodate these findings, I propose a *revamped ambiguity view*. Unlike prevailing pragmatic explanations, I contend that many dogwhistles exhibit semantic ambiguity between a widely accepted meaning and a secondary, controversial one. The dogwhistle

meaning arises from pertinent sub-communities developing new semantic conventions for word forms already used by other communities. Politicians, for instance, can exploit this phenomenon to garner support from fringe sub-communities without overtly committing to having meant what the dogwhistle means in the vocabulary of the relevant sub-community or to what that community represent (e.g., an ideology or some behaviors that are norm-violating or frowned upon). Theoretical support for this claim will be drawn from Lewis' (1975) notion of a *convention* and Clark's (1996) notion of a *communal lexicon*, which is a set of word-conventions specific to a sub-community.

Additionally, by adhering to empirical research on ambiguity, I will provide a detailed explanation of when dogwhistles are successful. The empirical research suggest that which reading the interpreter of ambiguous terms will go for is relative to the priming effects present in the context of utterance. I will use this to explain how dogwhistles' capacity to retain certain discourse effects is relative to the priming in the utterance contexts. It suggests that generally agreed upon features of dogwhistles, such as them being plausible deniable, will most likely only be present in an utterance context where both meanings are compatible with the context of utterance.

REFERENCES

Boholm, M. and Sayeed, A. (2023). Political dogwhistles and community divergence in semantic change. In Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change, pp. 53–65, Singapore. Association for Computational Linguistics.

Clark. H. H. (1996). Using Language. Cambridge University Press.

Gillon B.S. (2004). Ambiguity, indeterminacy, deixis, and vagueness: evidence and theory. In: Davis S, Gillon B.S. (eds) Semantics: a reader. Oxford University Press, Oxford: 157–190

Geeraerts. D. (1993). Vagueness's puzzles, polysemy's vagaries. Cogn Linguist 4(3): 223–272

Khoo, J. (2017). Code words in political discourse. Philosophical topics, 45(2): 33–64.

Lewis, D. (1969). Convention: A philosophical study. Harvard University Press.

Lowery, W. (2014). https://www.washingtonpost.com/news/the-fix/wp/2014/03/18/paul-ryanpoverty-dog-whistles-and-racism/ Accessed 3rd of February 2024.

Viebahn, E. (2018) Ambiguity and Zeugma. Pacific Philosophical Quarterly, 99: 749– 762.

The Response-Dependence of Meaning
C. Naomi Osorio-Kupferblum
University of Vienna, Austria | naomi.osorio@univie.ac.at

We can communicate with the help of language because language utterances mean something to us. In fact, an utterance is meaningful iff it means something to somebody. But being meaningful is a determinable property; the corresponding determinate properties just are the meaning each individual token utterance has. As these are still what the utterance means *to somebody*, I take such meaning to be a response-dependent property of that piece of language.

However, it is not entirely clear what a 'response-dependent property' metaphysically really is. Some consider them to be something like Lockean secondary qualities, as is the case for Johnston's 'response-dispositional' properties; some consider them rather like moral properties, and thus judgment-dependent – Crispin Wright, for instance, only distinguishes between properties detected and properties projected by *judgment*; Asta

sides with Euthyphro in this respect. Moreover, most have focused on concepts rather than the properties themselves. This includes prominently Pettit's account of global response-dependence – an account directly catering for aspects of language; and Nathaniel Goldberg's work on meaning. While Pettit's account was designed as an answer to the rule following problem and therefore includes an account of social practice, Goldberg analyses Davidson's account of triangulation and his swampman thought experiment. I will propose a truly response-dependent account – an account that regards response-dependent properties as only coming about in the interaction between the object to which we ascribe the property and the responder – as a further metaphysical option. I will then show how the distinction between 'response-dispositional', 'judgment-dependent', and 'response-dependent' properties corroborates Nathaniel Goldberg's analysis of two accounts of meaning in Davidson's work. I then argue that a truly response-dependent account of meaning explains best what sort of property the meaning of a verbal expression is.

Literature

Asta (Sveinsdóttir, Á. K.) (2010). Siding with Euthyphro: Response-Dependence and Conferred Properties. *European Journal of Philosophy*, *18*(1), 108-125.

Brynjarsdóttir, E. (2008) Response-Dependence of Concepts Is Not for Properties. *American Philosophical Quarterly* Vol. 45, 4, 377-386

Davidson, D. (2001). *Inquiries Into Truth and Interpretation: Philosophical Essays*. Clarendon Press.

Goldberg, N. (2009) Triangulation, Untranslatability, and Reconciliation. *Philosophia* 37, 261–280.

Goldberg, N. (2012). Swampman, response-dependence, and meaning. *Davidson's Philosophy: Truth, Meaning, and the Mental*, 148-63.

Johnston, M. (1989) "Dispositional Theories of Value", *Proceedings of the Aristotelian Society, Supplementary Volume* 63, 139-74.

Johnston, M. (1992). How to speak of the colours. *Philosophical Studies*, 68, 221–263.

Locke, J. 1979 (1689). *An Essay Concerning Human Understanding*, (ed.) P. H. Nidditch. New York: Oxford University Press.

Paseau, A. (2012). Against the Judgment-Dependence of Mathematics and Logic. *Erkenntnis* 76:23–40

Pettit, P. (1991). Realism and Response-Dependence. *Mind*

Pettit, P. (2005). *Rules, Reasons, and Norms: Selected Essays*. New York: Oxford University Press.

Wright, C. (1992). *Truth and objectivity*. Cambridge, MA: Harvard University Press.

Wednesday, 28/Aug/2024 9:40 – 10:20

Philosophy of Mind

Why Phenomenal Intentionality Theories Cannot Account for Identity Between Contents

Sara Papic

University of Milan, Italy | sara.papic@unimi.it

In the past decade, Phenomenal Intentionality Theories (PITs), which state that intentionality is ultimately grounded in phenomenal properties, have become a prominent alternative to "tracking" approaches to intentional content, such as causal or teleosemantic theories. I argue that strong versions of PIT cannot successfully account for identity between intentional contents. This is a problem because it is desirable for a theory of intentionality to be able to account for sameness and difference between the contents of intentional states.

To demonstrate that strong PITs cannot account for identity between contents, I begin by briefly laying out some of the minimal constraints of these theories when it comes to the

relationship between intentional content and phenomenal character. In particular, I express this through a fairly neutral supervenience thesis which states that, when it comes to at least some basic intentional states, there can be no difference in intentional content without there being a difference in phenomenal character. An additional preliminary note is necessary for my argument: introspection plays a key role in PITs, and specifically, introspective reports about the intentional content of experience are given great importance. Mendelovici, who is one of the most prominent proponents of a strong version of PIT, is explicit in this regard, and states that our theories of intentionality must match our pretheoretical introspective judgments about the contents of our experience to be adequate. In other words, PIT should methodologically prioritize introspection over other ways of determining the content of intentional states.

After having laid out these preliminary notes, I construct a scenario that shows how even these minimal constraints, along with the methodological priority given to introspection, lead to an unacceptable tension within PIT. The scenario is as follows: a subject faces a circular gradient composed of eight continuous shades (s_1-s_8). By design, s_1 is indiscriminable from s_2; s_2 is indiscriminable from s_3; and so on. However, s_1 is easily differentiated from s_8. I provide an image of the gradient to illustrate this.

Based on this scenario, one can lead the defender of PIT into contradiction. The subject will report that the content of their experience of s_1 and s_2 is identical; that the content of their experience of s_2 and s_3 is identical; and so on. Since identity is transitive, identity between contents should be transitive too, which means that we can infer that the content of the subject's experience of s_1 is identical to the content of their experience of s_8. However, the subject will report, based on introspection, that the content of their experience of s_1 and s_8 is *not* identical. We've stumbled into a contradiction.

The defender of PIT is faced with two unpleasant options here. One is to claim that one of the premises of the argument is false, that is, that one of the reports of the subject is false. However, this can only be done by abandoning the aforementioned priority of introspection, since all premises are supported by introspective reports. This is unacceptable because the prioritization of introspection is one of the fundamental ideas behind the program of phenomenal intentionality. The other option is to claim that the transitivity of identity does not hold for intentional contents, which is akin to saying that there can be no true identity between intentional contents, since transitivity is one of the defining characteristics of identity relations. This is unacceptable because the capacity to accommodate and explain sameness and difference of contents is a basic desideratum for any theory of intentionality.

I conclude this work by anticipating a possible counterargument. One might try to defend strong PITs by saying that there could be other factors determining content, and that PITs need not give total priority to introspection; for example, Mendelovici argues that our pretheoretical intuitions about content are sometimes related to the *psychological role* an intentional state plays (2018, pp. 27-8). I construct an argument to demonstrate why this is not a plausible way out for PIT. My argument has an analogous form to Mendelovici's "mismatch argument," which she employs to argue that tracking theories of intentional content are inadequate (2018, pp. 33-44) – it demonstrates that there could be a mismatch between the content determined by a state's psychological role and the content determined by a subject's introspective access to that same state. If such a mismatch occurs, I argue, the defender of PIT is constrained to give priority to what introspection dictates is the content of the intentional state. The argument can be generalized to target other kinds of potential theoretical "accessories" to PIT which would allow for alternative predictions about the

content of intentional states. Defenders of PIT, then, must prioritize introspection over alternative ways of determining the content of intentional states.

The upshot of this work is that strong versions of PIT, such as the ones developed in Mendelovici (2018) and Graham, Horgan and Tienson (2007), are committed to methodologically prioritizing introspection in a way that leads to them being unable to account for identity between contents in a satisfactory manner. I suspect this is a consequence of a deeper and more fundamental problem with strong PITs, which is that "flattening" representational properties onto phenomenal properties threatens to strip intentional states of their normative nature. In other words, strong versions of PIT such as Mendelovici's do not seem to leave space for intentional states to misrepresent; and there can be no representation without the possibility of misrepresentation. I believe more work needs to be done to investigate whether PITs can be modified to accommodate this basic feature of representation.

Bibliography:

Burge, Tyler (2007). *Foundations of mind*. New York: Oxford University Press.

Brewer, Bill (2011). *Perception and Its Objects*. Oxford, GB: Oxford University Press.

Campbell, John. (2002). *Reference and Consciousness*. Oxford: Oxford University Press.

Chalmers, David J. (2004). The representational character of experience. In Brian Leiter (ed.), The Future for Philosophy. Oxford University Press. pp. 153--181.

Dretske, Fred. (1995). *Naturalizing the Mind.* Cambridge: MIT Press.

Graham, George; Horgan, Terence E. & Tienson, John L. (2007). Consciousness and intentionality. In Max Velmans & Susan Schneider (eds.), *The Blackwell Companion to Consciousness*. Blackwell. pp. 468--484.

Horgan, Terence & Tienson, John (2002). The Intentionality of Phenomenology and the Phenomenology of Intentionality. In David J. Chalmers (ed.), Philosophy of Mind: Classical and Contemporary Readings. Oup Usa. pp. 520--533.

Mendelovici, Angela. (2018). *The Phenomenal Basis of Intentionality*. New York, USA: Oxford University Press.

Siewert, Charles P. (1998). *The Significance of Consciousness*. Princeton University Press.

Tye, Michael. (2000). *Consciousness, Color, and Content*. Cambridge: MIT Press.

Williamson, Timothy. (1994). *Vagueness.* New York: Routledge.

Yli-Vakkuri, Juhani, and John Hawthorne. (2018). *Narrow Content*, Oxford: Oxford University Press.

The Causal Exclusion Problem and James Woodward's Interventionism
Niccoló Aimone Pisano
University of St. Andrews, University of Stirling | nap3@stir.ac.uk

The causal exclusion problem has been one of the strongest and most debated metaphysical arguments against nonreductive physicalism in the philosophy of mind over the course of the past 25 years. The core idea of the problem is that, if one is committed to a rather minimal form of nonreductive physicalism, and if some plausible (although by no means uncontroversial) principles are granted, then one is forced to admit that higher-level properties, and mental properties in particular, do not possess causal powers.

Recently, there have been several notable attempts at offering solutions to this problem based on the adoption of Woodward's (2003) interventionist account of causation (e.g. List and Menzies (2010); Zhong (2014); Woodward (2015, 2022)), instead of directly rejecting one or more of the principles giving rise to the problem. However, it has also been argued (for

example by Baumgartner (2018)) that the exclusion problem cannot be dispelled as a result of the adoption of interventionism: this would only be the case if the supervenience bases of the relevant higher-level variables were independently fixable, which they are not for metaphysical reasons.

Since the interventionist arguments against the exclusion problem differ significantly from one another, the focus of my talk will be on just one of those interventionist solutions, namely the one proposed by Woodward (2015, 2022). This particular proposal revolves around an amended version of Woodward's original account of interventionism. The main difference between this "interventionism*" and the original version, consists in the fact that the latter is only concerned with purely causal graphs (i.e. with graphs only involving causal relations among variables), while the former explicitly takes into account causal graphs also involving metaphysical determination relations different from causation (in particular, supervenience relations). Specifically, interventionism* highlights the need *not* to hold the supervenience bases fixed while performing interventions on the relevant higher-level supervening variables, thus avoiding the aforementioned "independent fixability objection".

The main claim that I will argue for in my talk is that the reasons why one may be inclined to adopt interventionism* cannot be accepted in the context of the debate over the exclusion problem. To be clear, this is not to say that one may not legitimately adopt interventionism* for pragmatic purposes (hence in conformity with the general interventionist spirit) in ordinary causal reasoning. Rather, my claim is that, when it comes to the assessment of the exclusion problem, the considerations that play a role in deciding whether some variable is to be held fixed or not tilt the scale in favour of the need to hold the variables representing supervenience bases fixed, when an intervention is performed on the variables representing the relevant supervening (mental) properties. This is because failing to do so would make it impossible to answer the causal questions that motivate the discussion around the exclusion problem itself, specifically: "Do higher-level (mental) entities possess causal powers distinct from those of their supervenience bases?". But, as Woodward (2022, p.22) maintains, the causal questions one is interested in answering are part of the contextual factors guiding the decision to consider some variable as a confounder or not, and hence the choice of whether to hold it fixed or not. Therefore, failing to hold the supervenience bases fixed, as recommended by the amended version of interventionism, prevents one from ascertaining whether the causal powers that one may attribute to mental properties are also thereby possessed by the relevant supervenience bases, or whether they are indeed distinct.

My talk will be divided into two parts. In the first part, I will begin by offering a general presentation of the exclusion problem as a set of jointly inconsistent claims (*Nonreductionism*; *Physicalism*; *Closure*; *Exclusion*; *Causal efficacy*), as it has become more or less customary over time. After that, I will outline the core idea behind the interventionist account of causation: *X* is a direct cause of *Y* with respect to variable set V if there are possible interventions on *X* that will change the value of *Y* when all other variables in V are held fixed at some value by interventions (Woodward (2022, p.9). In doing so, I will explain how Woodward's recent amended version may be used to address the exclusion problem. Then, in the second part of my talk I will argue that this revised version of interventionism should not be accepted because it does not take into account the specific theoretical interests associated with the debate over the exclusion problem, as it instead should, based on Woodward's own recommendations.

References

Baumgartner, M. (2018). The inherent empirical underdetermination of mental causation. *Australasian Journal of Philosophy*, 96(2), 335-350.

Menzies, P., & List, C. (2010). The Causal Autonomy of the Special Sciences. In C. Macdonald, and G. Macdonald (eds), *Emergence in Mind*. Oxford: Mind Association Occasional Series.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, 91(2), 303-347.

Woodward, J. (2022). Modeling interventions in multi-level causal systems: supervenience, exclusion and underdetermination. *European Journal for Philosophy of Science*, 12(4), 59.

Zhong, L. (2014). Sophisticated exclusion and sophisticated causation. *The Journal of Philosophy*, 111(7), 341-360.

Wednesday, 28/Aug/2024 9:40 – 10:20

Philosophy of Language

Integrating Co-Speech Gestures into Sentence Meaning Comprehension

Ludmila Reimer[1], Mara Spychalska[2], Markus Werning[1]

[1]Ruhr-University Bochum, Germany; [2]Max Planck Institute for Psycholinguistics, Netherlands |
ludmila.reimer@rub.de

From everyday conversations, we know that gestures can enhance, disambiguate, or even change the intended meaning expressed by a speaker. Especially iconic co-speech gestures can provide nuanced information to, e.g., a talked about action, such as speed or orientation; and yet, gestures seem to not possess fixed meanings on their own. This begs the question whether and how humans use information present in gestures to process speech. To investigate whether iconic co-speech gestures modulate semantic predictions for upcoming linguistic content, we designed an EEG (electroencephalography) experiment. We focused on measuring the modulation of the N400 component on words that are expected or not, based on the gesture used in prior context, since the N400 is strongly linked to the predictability of the stimulus, e.g., it tends to be larger for words that are semantically less appropriate or less expected in a given context (Kutas & Federmeier, 2011).

We used naturalistic videos of a person saying a sentence uttering a general verb, featuring no gesture or an iconic co-speech gesture representing a more specific action. The target sentence was presented on screen and contained an instrument noun followed by its required action verb, both either matching or mismatching the previously seen gesture, giving us three conditions: Neutral (no gesture previously seen), Match, and Mismatch. After reading the target sentence, participants were prompted to rate whether the target sentence was a sensible continuation of the content of the context video. The linguistic material did not vary in modulating expectations about upcoming words, so any change during processing would be due to a change in the gesture. To avoid effects caused by unfitting gestures we ensured that all iconic co-speech gestures were congruent with the simultaneously uttered context sentence.

This approach is novel in several regards: First, experiments measuring ERPs elicited by linguistic targets following a gesture used paradigms of isolated, silent gesture videos followed by single probe words. Second, experiments using full sentences focused on the (mis)match between co-speech gestures and simultaneously uttered phrases or words (Hintz et al., 2023; Özyürek et al., 2007). Some also varied the linguistic context sentences, adding another factor potentially influencing the processing of utterances and co-speech gestures (Hintz et al., 2023). Nevertheless, these and other studies (see review by

Arachchige et al., 2021) showed that iconic co-speech gestures can interact with the semantic processing of linguistic input.

Since we consider iconic co-speech gestures as useful tools when it comes to disambiguating spoken utterances, we expected the Mismatches to be more surprising and thus elicit an N400 effect. For the Neutrals, our expectation was that we get an intermediate N400 effect since there is no gesture present that could lower or raise the expectation about upcoming input. Since we presented the target sentences word by word and the tool nouns preceded the action verbs, we additionally expect these effects to be stronger for the nouns as these were the first words either matching or mismatching the gesture.

For noun targets, we observed a significant negativity for Mismatch vs. Match that was in line with an N400 effect. For the verb targets, a significant negativity effect was found for the comparison Mismatch vs. Match, starting around 300ms and extending towards the end of the epoch, most prominent over the central-medial and posterior regions, suggesting a N400 effect with a sustained negativity. Another significant negativity effect is present in the comparison Match vs. Neutral, spanning over a similar time window.

For both targets, nouns and verbs, we found sustained positivity for Neutral-Match and Neutral-Mismatch which could be interpreted as a modulation of the P300, a positivity elicited between 250 and 500ms post-stimulus onset, known to be modulated by the probability of the target and by its relation to the task. Especially the P300b subcomponent tends to be larger for items that are task-relevant, and thus awaited in the experiment (Polich, 2003, 2007).

To conclude, listeners cannot only integrate the information presented in an iconic co-speech gesture, but the gestures also make a difference for a listener's probabilistic prediction regarding an upcoming instrument noun and a following action verb, and thus has a semantic effect on linguistic comprehension. The found N400 suggests that matching gestures facilitate linguistic understanding even for the upcoming linguistic content and not only for co-occurring linguistic content.

Alday, P. M., & Kretzschmar, F. (2019). Speed-Accuracy Tradeoffs in Brain and Behavior: Testing the Independence of P300 and N400 Related Processes in Behavioral Responses to Sentence Categorization. *Frontiers in Human Neuroscience*, *13*, 285. https://doi.org/10.3389/fnhum.2019.00285

Habets, B., Kita, S., Shao, Z., Özyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience*, *23*(8), 1845–1854. https://doi.org/10.1162/jocn.2010.21462

Hintz, F., Khoe, Y. H., Strauß, A., Psomakas, A. J. A., & Holler, J. (2023). Electrophysiological evidence for the enhancement of gesture-speech integration by linguistic predictability during multimodal discourse comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(2), 340–353. https://doi.org/10.3758/s13415-023-01074-8

Kandana Arachchige, K. G., Simoes Loureiro, I., Blekic, W., Rossignol, M., & Lefebvre, L. (2021). The Role of Iconic Gestures in Speech Comprehension: An Overview of Various Methodologies. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.634074

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605–616. https://doi.org/10.1162/jocn.2007.19.4.605

Polich, J. (2003). Theoretical overview of P3a and P3b. In Polich (Hrsg.), *Detection of Change: Event-Related Potential and fMRI Findings* (S. 83–98). Kluwer Academic Press: Boston.

Polich, J. (2007). Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, *19*(8), 1259–1274.

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language*, *101*(3), 234–245. https://doi.org/10.1016/j.bandl.2006.12.003

Monday, 26/Aug/2024 17:10 – 17:50

Philosophy of Mind

The Embodied Resonance in Episodic Memory through Re-enactment

Francesca Righetti

Ruhr University of Bochum, Germany | francesca.righetti@ruhr-uni-bochum.de

Contemporary theories of episodic memory skirt around an understanding of episodic memory as a mental scenario or simulation of a previous experience, advocating, to various degrees, for an internalist view. Traditionally, philosophers have assumed that remembering exclusively involves a "neural process, occurring inside an individual's brain" (Michaelian and Sutton, 2013). This view implies that memory entirely happens internally, and external resources are best understood as mere triggers. However, ever since Clark and Chalmers (1998) have notoriously challenged these philosophical implications, philosophers (such as Heersmink, 2017; Sutton, 2007) have suggested that remembering processes can rely on subjects' active involvement of embodied and embedded states or be scaffolded by bodily, environmental and social conditions.

It is not uncommon that subjects immerse themselves in previous bodily states to better recall, or, in other cases, to re-enact previous sequences of movements until they remember what they were doing or thinking. Empirical evidence has widely supported the idea that bodily procedures are involved in episodic memory, confirming two predictions: (i) behavioural re-enactment of processes involved in the encoding phase facilitates information retrieval; and (ii) accessibility to specific episodic memories may be blocked by a concurrent task involving the same sensorimotor resources (cfr. Ianí, 2019). However, investigations still need to clarify to what extent episodic memory can be embodied.

Through an analysis of the structures of subjective experiences of agency and recollections, this paper supports Ianí's claim that remembering an episode is intrinsically remembering actions (Ianí, 2019: 1762). I set out an investigation of embodiment in episodic memory on the subjective level of experience, from a phenomenologically-informed perspective. This allows us to understand the body not simply as the instantiation of sensorimotor patterns, but rather as the *lived* and *felt* body, and as the subjective *medium* for all our interactions with the world through our operative intentionality (generally understood as our practically oriented engagement with the environment).

Building upon the shared role of motor representations across procedural memory and episodic memory (Najenson, 2021), and following Sutton and Williamson's (2014) insights that there should not be any strict taxonomical separation between episodic and procedural memory, I provide an analysis of subjective experience of embodiment in episodic memory by comparing the role of operative intentionality and lived body in these two memory system. The analysis goes as follows.

Procedural memory concerns our habits and skills to engage with the world. The phenomenological analysis of the relationship between procedural memory and lived body

highlights two features: a flexible *re-enactment* of operative intentionality acquired over time and an *implicit awareness* of such knowledge. An expert tennis player *knows* how to hit the ball backhand, through the re-enactment of operative movements learned over time, without explicitly recollecting every single movement. Research on skill-based behaviour showed that "successful application (situation-dependent manifestation) of habits requires [...] at least minimal and sometimes intensified forms of awareness" (Tewes, 2018) and an implicit knowledge of bodily feelings. For example, when we perform difficult movements in sports, we *feel* an action was wrongly performed.

Episodic memory concerns personal past experiences. According to the traditional view of embodiment, any act of recollection includes a recollection of the whole past consciousness, including the implicit awareness of bodily past experiences (Thompson, 2010). Psychological literature emphasises the difficulty of sustaining that the evolutionary role of episodic memory is a faithful replay of our previous consciousness, it was rather developed to inform future behaviour (Suddendorf and Corballis, 2007). I argue that such a 'replay' of implicit consciousness is better explained as a flexible re-enactment of the previous operative intentionality. The practice of micro-phenomenological interviews (Petitmengin et al. 2019) supports the latter idea, by providing qualitative data on such a flexible re-enactment: during interviews, when subjects are induced to access their implicit and bodily awareness of past experiences, they involuntarily produce a series of embodied utterances, eyes movements to inspect absent scenarios, and hand gestures to represent actions.

However, such a re-enactment differs in aim from the one in procedural memory. Flexibly re-enacting the operative intentionality of previous experiences allows one to *embody* personal and others' agential perspectives, reliving and representing the past as actors in constructed scenarios (Trakas, 2018). On the other hand, present lived body and procedural abilities might affect episodic memory content, by impairing accessibility to episodic recollections when subjects find themselves in different bodily conditions and by feeling detached from past experiences (Ding and McCarroll, 2021).

Importantly, the experiential distance between the re-enacted episode and the recollecting act allows for the embodied resonance between the past and present. The concept of 'resonance' refers to two features: on one level, the offline re-activation of motor representations (i.e., motor resonance, which induces involuntary bodily feelings and muscle tension, cfr. Anelli et al., 2012); on another, the inherent otherness involved in episodic memory: there was an *agential I* that is no more.

Anelli, F., Borghi, A. M., & Nicoletti, R. (2012). Grasping the pain: motor resonance with dangerous affordances. Consciousness and Cognition, 21(4), 1627–1639.

Clark, A., & Chalmers, D. (1998). The Extended Mind. Analysis, 58(1), 7–19.

Dings, R., & McCarroll, C. J. (2022). The Complex Phenomenology of Episodic Memory: Felt Connections, Multimodal Perspectivity, and Multifaceted Selves. Journal of Consciousness Studies, 29(11–12), 29–55.

Heersmink, R. (2017). Distributed Selves: Personal Identity and Extended Memory Systems. Synthese, 194(8), 3135–3151.

Ianì, F. (2019). Embodied memories: Reviewing the role of the body in memory processes. Psychonomic Bulletin & Review, 26(6).

Michaelian, K., & Sutton, J. (2013). Distributed Cognition and Memory Research: History and Current Directions. Review of Philosophy and Psychology, 4(1), 1–24.

Najenson, J. (2021). Memory Systems and the Mnemic Character of Procedural Memory. The British Journal for the Philosophy of Science, 715448.

Petitmengin, C., Remillieux, A., & Valenzuela-Moguillansky, C. (2019). Discovering the structures of lived experience. Phenomenology and the Cognitive Sciences, 18(4), 691–730.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? Behavioral and Brain Sciences, 30(3), 299–313.

Sutton, J. (1998). Philosophy and memory traces: Descartes to connectionism. Cambridge University Press.

Sutton, J., & Williamson, K. (2014). Embodied remembering. In The Routledge handbook of embodied cognition (pp. 315–325). Routledge/Taylor & Francis Group.

Tewes, C. (2018). The Phenomenology of Habits: Integrating First-Person and Neuropsychological Studies of Memory. Frontiers in Psychology, 9.

Thompson, E. (2010). Self-No-Self? Memory and Reflexive Awareness. In M. Siderits, E. Thompson, & D. Zahavi (Eds.), Self, No Self?: Perspectives From Analytical, Phenomenological, and Indian Traditions. Oxford University Press.

Trakas, M. (2021). Kinetic Memories: An Embodied Form of Remembering the Personal Past. The Journal of Mind and Behavior, 42(2), 139–174.

Monday, 26/Aug/2024 16:30 – 17:10
Philosophy of Language

Quasi-Realist Approach to Normativity in Inferentialism

Szymon Sapalski
Jagiellonian University, Poland | szymon.sapalski@doctoral.uj.edu.pl

In this talk, I will present a quasi-realist approach to the problem of meaning normativity in inferentialism. Inferentialism is a metasemantic theory that explains the semantic content of expressions by relying on inferential relations between them and other expressions. One of the biggest problems for inferentialists is its potential incompatibility with naturalism. Put simply, inferentialists rely on a normative notion of rules to explain the semantic content of expressions which presents a challenge for naturalization, given the difficulty of naturalizing normativity. Recent literature has explored various solutions to the problem: showing that norms exist in the natural world (Hlobil 2022), appealing to the weaker forms of normativity (Glock 2022), or giving a causal story about emergence of rules (Peregrin 2022, Weiss 2022). In this talk, I will propose an alternative solution, drawing inspiration from Blackburn's (1993) quasi-realist approach to moral discourse and Reiland's account of regulative rules. This strategy will enable me to address two key objectives: 1) answering the metasemantic question of how expressions acquire their semantic content, and 2) offering a naturalistic explanation for why inference rules are normatively loaded.

Quasi-realism in metaethics acknowledges two perspectives on moral discourse—one external to the discourse from which naturalistic functional explanations of the discourse are provided and another, internal, from which one can talk about moral truths, properties, and beliefs. In a similar spirit I will distinguish two perspectives within an inferential linguistic practice. The explanation of what gives meaning to linguistic expressions comes from an external perspective. Expressions have meanings because a linguistic community accepts certain constitutive, naturalistically non-threatening inferential rules. Conversely, the internal-discursive perspective considers the viewpoint of participants in linguistic practice. From this standpoint, the semantic rules have their full normative force. However they also are not problematic from a naturalistic standpoint, as their normative character applies only when engaging in a language game.

To fully explain why expressions have meaning in inferentialist metasemantics, I will draw on Reiland's (2023) concept of regulative rules. This will also clarify both internal and

external perspectives on linguistic practice. According to Reiland, regulative rules are somewhere in between normative truths and orders. Unlike orders, they contain propositional content, and unlike normative truths, they can only be in force, not true. The acceptance of a rule within a community depends on their acceptance of what Reiland calls "bare propositional content." What I will suggest is that rules of inference should be understood precisely like Reiland's regulative rules. Unlike orders, each rule of inference carries specific propositional content. For example, the rule "From 'this is red' infer 'this is colored'" has a propositional content which can also be in force in a certain linguistic community. This second feature of inference rules (their ability to be in force) is what distinguishes them from normative truths.

The picture is, then, the following. What metaphysically constitutes an expression's meaning is a fact that members of a community follow specific inference rules (understood as regulative rules in Reiland's sense). I will argue that from an external perspective on linguistic practice this fact of following inferential rules can be explained solely in terms of dispositions. The fact that a rule is in force stems from the dispositions of individuals within a linguistic community to accept specific linguistic patterns. For instance, within a specific community, the phrase "This is red" means that something is red because individuals are disposed to accept an inferential move to "This is colored" but not to "This is blue." However, it may seem that there is more to rules being in force than merely individuals' dispositions toward certain linguistic behaviors. There appears to be a sense of normative force linked to the acceptance of rules. I will argue that this appearance of normativity is captured by considering the internal point of view of linguistic practice. The internal perspective is characterized by participants of linguistic practice having a pro-attitude toward specific propositional content. This pro-attitude not only establishes that a particular rule is in force within a linguistic community but also gives it a normative force. In contrast, from an external standpoint, one can only hold a belief, formed by observing the dispositions of a particular community, that a linguistic rule is in force. Importantly, this belief does not fully capture what it means to follow a particular rule; namely, it does not grasp the normative, inherently motivating aspect of the rule.

In a nutshell, the quasi-realist view I propose offers a naturalistic account of what determines the semantic content of expressions: people's dispositions to follow regulative inferential rules. The account I am defending also explains why these rules generate normative force—it is because members of the linguistic community hold a pro-attitude towards them.

References:

Blackburn, S. (1993) Essays in Quasi-Realism. New York: Oxford University Press.

Glock, H. J. (2022) "Norms, Reasons, and Anthropological Naturalism." Philosophical Topics, Vol. 50, No. 1: 9–32.

Hlobil, U. (2022) "Teleo-Inferentialism." Philosophical Topics, Vol. 50, No. 1: 185–211.

Peregrin, J. (2022) "Inferentialism Naturalised." Philosophical Topics, Vol. 50, No. 1: 33–54.

Reiland, I. (2023) "Regulative Rules: A Distinctive Normative Kind." Philosophy and Phenomenological Research.

Weiss, B. (2022) "From Tools to Rules: The Evolution of Rule-Following." Philosophical Topics, Vol. 50, No. 1: 55–82.

Common Ground beyond the Grave

Merel Semeijn

École Normale Supérieure, France | semeijn.merel@gmail.com

As recently described by Geurts (2023), the study of the notion of 'common ground' in pragmatics takes face-to-face conversations as a model for communication. In face-to-face settings, interlocutors typically have (more or less) *simultaneous* attitudes concerning conversational updates: the moment that the speaker states that *p*, the addressee will immediately also come to believe/accept that *p* (or at least that *p* was stated). Likewise, formal characterizations of common ground on offer, (e.g., in terms of common belief/ acceptance/knowledge/commitment, see e.g., Stalnaker (2014); Clark and Marshall (1981); Geurts (2019)) typically assume *simultaneous* attitudes. At best, these definitions do not say anything about *when* interlocutors have the relevant attitudes.

However, the concept of common ground has, without much hesitation, (except by Harris (2020); Semeijn (2024)) been extended to asynchronous non-face-to-face conversations in which speaker and addressee do *not* have simultaneous attitudes. For instance, several authors in philosophy/semantics of fiction (e.g., Zucchi (2021); Eckardt (2014); Maier and Semeijn (2021)) use the notion of common ground between an author (e.g., Austen) and her audience (e.g., myself) to analyze fictional discourse and/or truth in fiction.

The intuitive justification for this move is simple: Non-face-to-face conversations work in essentially the same way as face-to-face conversations. Consider a simple case of asynchronous communication: A man on his deathbed writes a letter to his daughter that states "The medallion is in Nouvion" (*p*) and dies. The daughter reads her father's 'letter from beyond the grave' and comes to believe that *p*. Similar to cases of face-to-face communication, the notion of 'common ground' has a dual function (see Geurts (2023)) to play in our explanation of this communicative act: First, in writing his letter, father proposed to update the common ground between him and his daughter with *p*. Second, father was able to felicitously phrase his contribution in this way (i.e., using the definite description "the medallion") because it was already common ground between him and his daughter that there was some specific medallion.

As natural as that may sound, if we assume straightforward time-indexing (i.e., the kind of time-indexing that makes perfect sense for face-to-face settings) for standard 'mentalistic' characterizations of common ground (e.g., in terms of common belief), then *p* is *never* actually common ground between father and daughter. Below $B_{x\,n}^{t}\phi$ means *x* believes that ϕ at *tn*:

p is common ground at *t1* between speaker *a* and hearer *b* iff

$B_{a\,1}^{t}p$
$B_{a\,1}^{t}B_{b\,1}^{t}p$
$B_{a\,1}^{t}B_{b\,1}^{t}B_{a\,1}^{t}p$
...
$B_{b}^{t1}p$
$B_{b\,1}^{t}B_{a\,1}^{t}p$
$B_{b\,1}^{t}B_{a\,1}^{t}B_{b\,1}^{t}p$
...

By the time that the daughter comes to believe that *p*, the father has already passed away.

Multi-agent system logics may aid philosophers of language here: Loosely following Halpern and Moses's (1990) concept of 'eventual common knowledge', we might propose the following definition of 'eventual common ground':

p is common ground between speaker a and hearer b iff

$\exists t B_a^t p$

$\exists t B_a^t \exists t' B_b^{t'} p$

$\exists t B_a^t \exists t' B_b^{t'} \exists t'' B_a^{t''} p$

...

$\exists t B_b^t p$

$\exists t B_b^t \exists t' B_a^{t'} p$

$\exists t B_b^t \exists t' B_a^{t'} \exists t'' B_b^{t''} p$

...

Now p *is* common ground between father and daughter: They both believe that p at some point in time, both believe at some point in time that the other believes that p at some (possibly other) point in time, etc.). However, this notion also will not do. On this definition it is *always* common ground between father and daughter that p. However, we require a notion of common ground that is *dynamic*: Common ground is something that can grow as communication progresses. Intuitively, p *became* common ground at some point thanks to father's letter.

I propose that we require a shift from defining 'common ground between agents' at a certain time to defining common ground between 'agents at a certain time':

p is common ground between speaker a at t_1 and hearer b at t_2 iff:

$B_a^{t_1} p$

$B_a^{t_1} \exists t (t \geq t_1 \wedge B_b^t p)$

$B_a^{t_1} \exists t (t \geq t_1 \wedge B_b^t \exists t' (t' \leq t \wedge B_a^{t'} p))$

...

$B_b^{t_1} p$

$B_b^{t_1} \exists t (t \geq t_1 \wedge B_a^t p)$

$B_b^{t_1} \exists t (t \geq t_1 \wedge B_a^t \exists t' (t' \leq t \wedge B_b^{t'} p))$

...

The basic idea is that in case father had the relevant beliefs in *his time* (e.g., at t_1 father believes that p and that his daughter would come to believe that p, etc.) and the daughter had the relevant beliefs in *her time* (e.g., at t_2 daughter believes that p and that her father used to believe that p), then it is common ground that p between father *in his time* and daughter *in her time*. I explore predictions of and potential issues with this definition (e.g., this definition allows us to specify what is common ground between speaker at t_2 and hearer at some *earlier* time t_1, a situation arguably only possible in fiction).

References

Clark, H. H. and Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A. K., Webber, B. L., and Sag, I. A., editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge University Press, Cambridge.

Eckardt, R. (2014). *The Semantics of Free Indirect Discourse*. Brill Publishers, Leiden.

Geurts, B. (2019). Communication as commitment sharing: speech acts, implicatures, common ground. *Theoretical Linguistics*, 45:1–30.

Geurts, B. (2023). Common ground in pragmatics.

Halpern, J. Y. and Moses, Y. (1990). Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587.

Harris, D. W. (2020). We talk to people, not contexts. *Philosophical Studies*, 177(9):2713–2733.

Maier, E. and Semeijn, M. (2021). Extracting fictional truth from unreliable sources. In Maier, E. and Stokke, A., editors, *The Language of Fiction*, pages 186–220. Oxford University Press, Oxford.

Semeijn, M. (2024). Common ground: *In sensu composito* or *in sensu diviso*. *Journal of Philosophical Logic*.

Stalnaker, R. (2014). *Context*. Oxford University Press, Oxford.

Zucchi, A. (2021). On the generation of content. In Maier, E. and Stokke, A., editors, *The Language of Fiction*, pages 107–130. Oxford University Press, Oxford.

Monday, 26/Aug/2024 12:10 – 12:50
Philosophy of Mind

Self-Ascription and Primitiveness

Daniel Skibra
University of Konstanz, Germany | danielskibra@gmail.com

Outline.
In this talk, I do three things. First, I examine a kind of objection to the account of the attitudes given in Lewis 1979 which criticizes the notion of self-ascription as a putative unexplained primitive of the theory. I argue that it's not so much primitiveness *per se* that is objectionable, but diagnose two consequences of this primitiveness which are objectionable. (Cf. Stanley 2011; Capellen and Dever 2013; Holton 2015) Second, I consider recent defenses of Lewis's theory (due to Jackson and Stoljar 2020; Openshaw 2020) which contend that self-ascription is not primitive and offer explanatory strategies. In light of my diagnosis, I argue that the more important question is not whether self-ascription is a primitive, but whether these strategies can meet the objections I diagnose. My take on this question is salutary—these strategies can defuse much of the objections. Third, I argue that there is still a lingering problem regarding self-ascription which these explanatory strategies bring into relief.

Lewis's property theory.
In Lewis 1979, Lewis advocates a novel theory of the attitudes motivated in part by the fact that the more traditional, propositionalist, possible worlds theory of the attitudes he previously endorsed was unable to represent the kinds of de se attitudes discussed by Castaneda (1966), Perry (1979), etc., nor to distinguish them in a satisfactory way from their de dicto counterparts. Lewis's proposal was novel in two ways. First, it allowed attitudes to be relations to properties instead of propositions, construed as sets of centered worlds (formally, world-center pairs, understood as world-bound time-slices of individuals). Second, it construed the relation a subject has to content in terms of self-ascription. Taken jointly, this allows Lewis's theory to capture both de se and de dicto attitudes in a general theory of the attitudes. Lewis also leveraged this framework for an account of de re attitudes, which I will not address directly in this talk.

First: primitive self-ascription—what's the objection? While Lewis's theory has proved tremen- dously influential, and the centered worlds framework based on it very productive, a number of objections to Lewis's theory have emerged in prominent venues in the last decade. These objections are motivated by a common complaint—that the self-ascription relation, which is fundamental to Lewis's theory and provides the linch-pin for the uniform account, is an unexplained primitive. This primitiveness is unmotivated and leaves the whole theory on shaky footing. In examining this criticism, I find that it is not the apparent primitivism that is problematic, but some consequences that flow from it. We do better, I argue, to understand the primitiveness complaint in terms of two kinds of objections. The first (of which Stanley 2011 is a representative) contends that Lewis's account evades a

description of what cognitive contact to the self amounts to. The second (which I associate with Capellen and Dever 2013) has to do with a kind of arbitrariness at the heart of self-ascription. This objection contends that there is no basis for the central technical innovation of Lewis's account.

Second: do the explanations defuse the objections? Recent defenses of Lewis (e.g., by Jackson and Stoljar (2020); Openshaw (2020)) have appeared, objecting to the characterization of self- ascription as a primitive. Their main aim is to show how self-ascription is explained by other resources in Lewis's work. The explanatory strategy in Jackson and Stoljar 2020 appeals to Lewis's functionalist commitments. The strategy in Openshaw 2020, to some of Lewis's foundational remarks on semantics, drawing on the concept of initialization. (The term is not Lewis's but can be given a Lewisian gloss.) I argue that, in light of the diagnosis of the primitiveness complaint above, simply pointing out that self-ascription is not primitive is insufficient. We need to see how the explanation for self-ascription can deal with these deeper objections. Though the proposals on offer are different, they are actually complementary. Jackson and Stoljar's functionalist strategy can help deal with the evasion objection, and Openshaw's initialization strategy helps deal with the arbitrariness objection. This is a welcome result for the friend of the property theory.

Third: a remaining wrinkle for self-ascription I then argue a final point. Notwithstanding the promising explanatory strategies, there is a remaining wrinkle when it comes to the objections to self-ascription. The initialization strategy provides a pretty promising response to the arbitrariness objection, but it makes an empirical prediction about the nature of the attitudes and attitude content. It commits one to what Recanati (2007) calls the "reflexive constraint". (Cf. also Pagin 2016 for a similar point.) In terms of the initialization strategy, it is the commitment that the initializing indices, (Lewis's "original indices"; Lewis 1983) are those of the context of utterance/ thought. Appeal to the semantics of natural language can make the reflexive constraint seem plausible, but the move is questionable when applied to the attitudes. Recanati gives grounds for questioning it unrelated to the dispute about self-ascription. I offer some additional reasons: there are non-doxastic attitudes that have contents not plausibly described being bound by the reflexive constraint. I conclude by reflecting on where this leaves the explanation of self-ascription vis-a`-vis the aforementioned objections.

References

Capellen, H. and J. Dever (2013). The Inessential Indexical. OUP.

Castaneda, H. (1966). 'He': A study in the logic of self-consciousness. Ratio 8, 130–157.

Holton, R. (2015). Primitive self-ascription: Lewis on the de se. In B. Loewer and J. Schaffer (Eds.), The Blackwell Companion to David Lewis. Blackwell.

Jackson, F. and D. Stoljar (2020). Understanding self-ascription. Mind and Language 35(2), 141–155.

Lewis, D. (1979). Attitudes de dicto and de se. Philosophical Review 88(4), 513–543.

Lewis, D. (1983). Postscripts to General Semantics. In Philosophical Papers, Volume 1, pp. 230–232. OUP.

Openshaw, J. (2020). Self-ascription and the de se. Synthese 197(5), 2039–2050.

Pagin, P. (2016). De se communication : Centered or uncentered? In M. Garćıa-Carpintero and

S. Torre (Eds.), About Oneself: De Se Thought and Communication. OUP.

Perry, J. (1979). The problem of the essential indexical. Nouˆs 13(December), 3–21.

Recanati, F. (2007). Perspectival Thought. OUP.

Stanley, J. (2011). Know how. OUP.

Non-Assertoric Moorean Speech Acts

Maciej Tarnowski

University of Warsaw, Poland | m.tarnowski3@student.uw.edu.pl

An orthodox characteristic of Moore's Paradox (Moore 1942) presents it as a problem of adequately explaining why the following *assertions* seem self-contradictory or absurd, although their content is perfectly consistent:

(MP) *p,* but I don't believe that *p*.

Moore's Paradox is therefore often dealt with by appealing to properties of the speech act of assertion and the belief it expresses. Many theorists (e.g. Black 1952, Searle 1970, Shoemaker 1995, Alston 2000, Woods 2018, Mandelkern 2021) noted, however, that Moore-like phenomena also seem to arise in cases of (partly) non-assertoric speech acts, such as:

(1) Please *phi*, but I don't want you to *phi*.

(2) *p*? - but I don't want to know if *p*.

(3) I order you to *phi*, but you may not *phi*.

(4) I promise to *phi*, but I will not *phi*.

This observation according to some (Woods 2018) stands against the standard analysis of Moore's Paradox in terms of expressing irrational belief. In my talk, I will aim at explaining the puzzling nature of such sentences without dropping this crucial assumption and demonstrate, that such constructions reveal interesting, yet widely neglected features of non-assertoric speech acts and their interaction with the conversational common ground.

I will start by presenting and critiquing the characterization of these speech acts present in the literature, distinguishing different forms of Moorean speech acts, which I label "sincerity" (i.e. denying, in the second conjunct, the sincerity condition of a speech act performed in the first conjunct, e.g. (1)-(2)) and "satisfaction" formulations (denying, in the second conjunct, the satisfaction condition of the mental state expressed in the first conjunct, e.g. (3)-(4)). I shall argue that blurring this distinction led to many existent solutions to lack needed generality: in particular Searle's (1969) and Woods' (2018) analyses in terms of speaker's commitments and Mandelkern's (2021) analysis in terms of conversational pretense. In contrast, I will hypothesize that non-assertoric speech acts have a double effect on the common ground (understood as common belief or common knowledge) presupposed by the speakers usually thought to be reserved for assertions (Stalnaker 1978, 1984): that, once accepted, such speech acts make speakers accept that the speaker is in a mental state conventionally expressed by these acts and that their satisfaction condition is or will be met:

(B) By performing the speech act *α,* conventionally expressing mental state *M* with a satisfaction condition *s*, the speaker proposes to add the proposition that *s* occurs or will occur to the common ground (e.g., by promising that they will *phi*, the speaker to add to the common ground the proposition that *they will phi*).

(R) By performing the speech act *α,* conventionally expressing mental state *M* with a satisfaction condition *s*, the speaker offers *M* as a reason for why *s* will obtain, obtains, or why the hearer should make it the case (e.g., by promising that they will *phi*, the speaker offers their intention to *phi* as a reason why the proposition that *they will phi* is true).

B and R both provide the immediate solution to the initial puzzle. Given that non-assertoric Moorean speech acts conjoin both (a) the proposal to add to the common ground the proposition that *s* obtains or will obtain or that they are in *M*, and (b) the proposal to add to the common ground the proposition that *s* does not or will not obtain or that they are not

in *M*, in effect they aim at introducing contradictory propositions to the common ground, which gives rise to Moore-paradoxicality.

In the remainder of my talk, I shall discuss motivations standing behind accepting both B and R as descriptive claims about the effects of non-assertoric speech acts on the common ground and some possible objections towards it. This claim builds on two crucial observations: (a) the fact that non-assertoric speech acts introduce new intuitively acceptable presuppositions to conversation, (b) that non-assertoric speech acts may often be realized by assertoric means reflecting the content of B and R (e.g. one may promise by simply asserting "I will phi" or "I intend to phi"). Building upon the observations of Ninan (2005) and Portner (2004, 2007), I will claim that these two features may be fruitfully explained in the common ground framework by appealing to the fact that non-assertoric speech acts may serve as a way of establishing a common plan and allow for rational coordination.

References:

Alston, W. P. (2000). *Illocutionary acts and sentence meaning*. Ithaca: Cornell University Press.

Black, M. (1952). Saying and disbelieving. *Analysis*, *13*(2), 25-33.

Mandelkern, M. (2021). Practical Moore sentences. *Noûs*, *55*(1), 39-61.

Moore, G. E. (1942). *Reply to My Critics*. In P. A. Schlipp (Ed.), *The Philosophy of George Edward Moore*, Menasha: George Banta Publishing Company.

Ninan, D. (2005). Two puzzles about deontic necessity. In *New Work on Modality. MIT Working Papers in Linguistics*, 51 (2005).

Portner, P. (2004). The semantics of imperatives within a theory of clause types. In *Semantics and linguistic theory* (Vol. 14, pp. 235-252).

Portner, P. (2007). Imperatives and modals. *Natural language semantics*, *15*, 351-383.

Searle, J. (1970). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Shoemaker, S. (1995). Moore's paradox and self-knowledge. *Philosophical Studies, 77*(2/3), 211-228.

Stalnaker, R. C. (1978). Assertion. In *Pragmatics* (pp. 315-332). Brill.

Stalnaker, R. C. (1984). *Inquiry.* Cambridge: MIT Press.

Woods, J. (2018). A commitment-theoretic account of Moore's paradox. In *Pragmatics, truth and underspecification* (pp. 323-348). Brill.

Monday, 26/Aug/2024 17:50 – 18:30
Philosophy of Language
An Experimental Approach to Empty Definite Descriptions
Renato Turco

University of Genoa, Italy | r.turco91@gmail.com

Empty definite descriptions have posed a challenge for any philosopher who engaged with Russell's seminal "On Denoting". Russell held the view that sentences containing them are always false, while Strawson argued that such sentences have no truth-value at all, for a *presupposition failure* occurs. However, he added, the observation that a statement like "The present King of France is bald" is truth-valueless is based on *intuition* (cfr., Strawson 1950, 1964). This situation serves as a starting point for engaging with the issue with the help of *experimental linguistics*.

A test on definite descriptions has been already carried out a decade ago by Márta Abrusán and Kriszta Szendröi; the purpose of their experiment was to "investigate the truth value intuitions reported in the literature" (Abrusán and Szendröi 2013, p. 2) of empty

definite descriptions. The experimenters have analyzed how - and if - some pragmatic factors influence speakers' truth-value judgments about such sentences: these factors are the presence of a background knowledge (Lasersohn 1993), verifiability (Von Fintel 2004), and topicality (Strawson 1964). Their result is that "verifiability seems to be the determining factor in native speakers' judgments of sentences involving reference failure of noun phrases. Topicality is parasitic on that." (*ibidem*, p.35).

However, the test carried out by the authors presents some serious flaws:

– The total number of conditions in this experiment is 11, each consisting in 8 critical items (8 sentences containing empty descriptions), alongside twice as many fillers. This amounts to a rough total number of 250 sentences, administered to 33 participants. This disproportion is statistically and methodologically troublesome.

– Strawson's topicality has been triggered by employing *left dislocated structures*, i.e., constructions like "The King of France, he was invited to have dinner with Sarkozy". If the effects observed are due to the topical status of the definite description, the same effects should be obtained in other topical constructions like *clefts*.

– Employing a *ternary judgment task* (the participant had to choose among TRUE, FALSE, CAN'T SAY) is debatable. Apart from considerations about the lexical meaning of the label the authors have decided to use, asking directly to participants to adjudicate between three values means that they are made aware of the fact that there is, indeed, a third value to begin with, that may be ascribed; the distinction between falsity (FALSE) and presupposition failure (CAN'T SAY) could be, in fact, not available for non-experts.

Therefore, I carried out an experiment so to amend the three points I have mentioned above. In a within-subject design, 68 participants, divided into two sets, had to perform a rating task: they had to rate, on a Likert scale from 1 to 7 (1 being "False" and 7 being "True"), how much they were inclined to ascribe truth to the sentences they were reading. I have compared two Strawson's conditions: a presuppositional condition (e.g., "The King of France is bald" vs. "France has a King, and he is bald") and a topicality condition (e.g., "It is the King of France who called Macron last night" vs. "It is Macron who called the King of France last night"). Each set consisted in a number of 66 items in a semi-randomized order.

With respect to the presuppositional condition, cross-analysis on the results shows that, when the existential presupposition is made explicit, speakers tend to evaluate the sentences more falsely than cases in which existence is presupposed (71,43% vs. 57,58%). I believe that these results are consistent with Strawson's hypothesis, in the fact that speakers *use* the asserted existence claim to evaluate – falsely – the statement. When the presupposition is not asserted, subjects face some difficulties in the evaluation, and this is mirrored by the percentage of responses falling around the mid-value (27,27%).

With respect to topicality, the results in Abrusán and Szendröi (2013) are not confirmed: when the referentially sound NP (in my example, "Macron") is the clefted constituent, more false outcomes are observed than cases in which the empty description is moved higher in the sentence (75,76% vs 65,71%). Observe that the clefted constituent is *not* the topic expression, in an aboutness sense; rather, it is the part of the sentence occurring after the complementizer.

In Abrusán and Szendröi (2013), left dislocated structures are employed, and it is argued that their results are expected also if different syntactic environments triggering topicality are used. The fact that the results obtained point towards another direction calls, at least, for further investigations, both from an experimental point of view and from a theoretical perspective, most notably on topics and cleft constructions, which carry nontrivial considerations on pre- supposition accommodation (Ebert & Ebert 2010; Felka 2015). Moreover, these findings should be supported by supplementary studies, employing online

and not offline measures. However, I believe that the present results are suitable as a starting point, for, at the very least, the proportion of items per participants is much more statistically stable.

References

Abrusán, Márta and Kriszta Szendröi (2013). "Experimenting with the King of France: Topics, Verifiability and Definite Descriptions". In: *Semantics and Pragmatics* 6.10, pp. 1–43.

Ebert, Christian and Cornelia Ebert (2010). "On squeamishness of the royal kind". In: *Language and Logos: Studies in Theoretical and Computational Linguistics* 72, p. 149.

Felka, Katharina (2015). "On the presuppositions of number sentences". In: *Synthese* 192.5, pp. 1393–1412.

Lasersohn, Peter (1993). "Existence presuppositions and background knowledge". In: *Journal of semantics* 10.2, pp. 113–122.

Russell, Bertrand (1905). "On Denoting". In: *Mind* 14.56, pp. 479–493. Strawson, Peter F. (1950). "On Referring". In: *Mind* 59.235, pp. 320–344.

— (1964). "Identifying Reference and Truth-Values". In: *Theoria* 30.2, pp. 96– 118.

Von Fintel, Kai (2004). "Would you believe it? The King of France is back! (Pre- suppositions and truth-value intuitions)". In: *Descriptions and Beyond*. Ed. by Marga Reimer and Anne Bezuidenhout. Oxford University Press, pp. 315– 341.

Thought, Indexicals, and Composite MOPs

Víctor Martin Verdejo
Pompeu Fabra University, Spain | vmverdejo@gmail.com

With roots in Frege's famous remarks (1956, 296), reflection on Rip van Winkle's fantastic story has played a key role in the philosophical study of indexical or cognitive dynamics (Kaplan 1989, Perry 1997, Branquinho 2008, Ludlow 2019). Consider now the Reverse van Winkle case: Rip van Winkle falls asleep at time t and, while he feels like it's been a very long slumber, only a few seconds have actually passed when he wakes up at time t'. Suppose Rip van Winkle utters "Today is fine" both at t and t' but, while he fully accepts the associated thought at t, he hesitates at t'.

The Reverse van Winkle and structurally similar cases show that, if we accept the 'Intuitive Criterion of Difference' (Evans 1982), a particular understanding of the relation between indexicality and thought is wrongheaded. According to this 'linguistic' view, indexicality is a property of linguistic terms only and these terms express thoughts relative to a particular context. If this view were correct, sameness of context of utterance, indexical expression and reference should guarantee sameness of thought. However, the target case shows that the *same* (day-based) indexical term –"today"– and the *same* relevant context to refer to the *same* day may involve conflicting rational attitudes, and hence *different* thoughts. The case can be raised even if one doesn't accept (contra Perry 1997: 35-38 or Ludlow 2019: 72-75) that the first "today"-thought at t is remembered at t'. One only requires that Rip, at t', doesn't change his mind with respect to the thought expressed at t (cf. Kaplan 1989: 537-538).

We should not however haste to embrace the view that indexicality is itself an essential aspect of thought. If this were so, it should be possible for thoughts to be indexically individuated. Yet sometimes, as the (Reverse) van Winkle case illustrates, thoughts expressed with (same or different) indexicals change with contexts and sometimes they

don't. This suggests that there are no straightforward contextual factors responsible for the individuation of thought of the sort one would expect if thought were essentially indexical.

What should be done? To analyse the target and related cases, I will invoke the notion of a "composite mode of presentation", i.e. "a mode of presentation that, although 'static', i.e. deployed at a given time in thinking of the object, rests on *distinct simultaneous relations* to the object, and on distinct ways of gaining information (distinct information channels) based on these relations" (Recanati, forthcoming; see also Dickie & Rattan 2010, Recanati 2016).

Thus, in the Reverse van Winkle scenario, at t', Rip takes recourse to two different modes of presentation (MOPs) of a particular day, one based on memory awareness of it before falling asleep, and one based on his direct awareness of the day in question. While one would typically merge these MOPs into one composite MOP to think and reason, indexically, about a day, Rip van Winkle fails to do so because of his especial predicament. Rip van Winkle has different thoughts, based on different MOPs. But these MOPs would typically constitute one and the same composite MOP in normal circumstances.

More needs to be said, however, to fully characterize the cases in which composite MOPs based on different primitive MOPs of a referent are indeed available. My proposal is that this happens when the thinker is aware of the co-referentiality of primitive MOPs. Such awareness – which can be spelled out in a number of ways – may link together very different indexical and demonstrative MOPs (perceptual, testimonial, memory-based…). However, composite MOPs need not be restricted to thought expressible with indexicals or demonstratives, and may carry over to any co-referential singular and general terms. This suggests a view in which the MOPs associated with indexicals are correctly attributed to the thought itself, but where such MOPs are not actually so different from conventional, non-indexical MOPs. Finally, while the awareness of co-reference signals the presence of a composite MOP, there is a sense in which composite MOPs may be acknowledged whether or not a thinker – such as Rip van Winkel – is aware of the co-referentiality of their constituent MOPs. This is also the sense in which different subjects unaware of one another may express the very same thought via different utterances of "Today is fine" on the same day. This point illustrates the mind-independent character of composite MOPs.

References

Branquinho, J. (2008) On the Persistence and Re-Expression of Indexical Belief. *Manuscrito* 31: 573-600.

Dickie, I. and Rattan, G. (2010) Sense, Communication, and Rational Engagement. *Dialectica* 64: 131-51.

Evans, G. (1982) *The Varieties of Reference* (edited by J. McDowell). Oxford: Clarendon Press.

Frege, G. (1956) The Thought. *Mind* 65: 289-311.

Kaplan, D. (1989) Demonstratives. In J. Almog, H. Wettstein and J. Perry (eds.), *Themes from Kaplan*, 481-563. New York: Oxford University Press.

Ludlow, P. (2019) *Interperspectival Content*. Oxford: Oxford University Press.

Perry, J. (1997) Rip Van Winkle and Other Characters. *European Review of Philosophy* 2: 13-39.

Recanati, F. (2016) *Mental Files in Flux*. Oxford: Oxford University Press.

Recanati, F. (forthcoming) Sameness of Mode of Presentation. In Bermúdez, J.L. et al. (eds.) *Sharing Thoughts*. Oxford: Oxford University Press.

Problems with Propositional Attitude Accounts of Imagination: A need to Account for the Imaginative through Flexibility

Alyssa Suzanne Walker

Tulane University, United States of America | awalker11@tulane.edu

Propositional attitude accounts of imagination (PA-I) currently fall short of explaining an important aspect of imagination. They do not explain why we attribute imaginativeness to a person when they use their imagination, nor do they explain that imaginativeness can come in degrees. We often attribute imaginers as being imaginative when they engage in pretense, fiction, and the creation of ideas and achievements. I provide cases that show that being imaginative picks out more than just the fact that one mentally represented some content that is neither perceptually available nor truth-directed; it refers to the particularity of how and what someone imagines. I argue that propositional attitude accounts can make sense of this by supplementing the view with a concept of flexibility.

A representative PA-I account comes from Nichols and Stich (2000), for whom imagination, conceived of as the representation of i-beliefs, is a propositional attitude that represents information as it would be given certain assumptions. The goal of their account is to explain how the imagination fits into our mental economy among other folk psychological states. Nichols and Stich offer i-beliefs to explain how we produce pretense behavior. Since we often attribute imaginativeness to pretenders, we should expect an account like this to explain what grounds such attributions.

I demonstrate that PA-I accounts fail to explain what makes someone imaginative. The representation of an i-belief is insufficient because it does not explain how one comes to have the imagining they do, and this is what is picked out by the concept of imaginative. PA-I only describes the resultant state as one of imagination, whereas we ought to describe the process that underlies it as being imaginative. A similar point was made by Wiltsher (2022) and Stokes (2014, 2016), who argue for a view of imagination as a process or as the ability to manipulate, respectively. However, neither Wiltsher nor Stokes describe the kind of process or manipulation required to do this job.

The concept of flexibility can provide us with the answer. We are attributed as being imaginative when we are able to control how we move information among settings. Cognitive scientists and psychologists describe this as 'cognitive flexibility', a type of executive function. They use flexibility to explain creativity (for example, Zabelina et al. 2010, 2018), but it can also explain the range of behaviors to which imaginativeness is attributed (including non-creative ones). Flexibility explains that we can organize information based on different principles, and that such organization requires moving information among settings. Moreover, one can organize information in highly flexible ways or in less flexible ways.

I look at two experiments in psychology that have demonstrated a relationship between imagination and flexibility. I argue their findings support supplementing accounts of imagination with a concept of flexibility. Veraksa et al. (2022) hypothesized that if one demonstrated a higher development of imagination, then one has a higher development of executive functioning. The authors asked 206 Russian school-aged children to complete ten drawings. They presented the students with cards, on which was a partial drawing. Students were asked to create a "whole" drawing, with no time limit, but were required not to erase or change anything once they started. They measured imagination development in terms of uniqueness, detail, and strategy, and compared these measures to executive functioning

through card sorting tasks. The authors used general linear models to demonstrate that higher degrees of imagination co-occurred with higher degrees of executive functioning, in particular cognitive flexibility.

White (2018) compared 52 American college students with and without diagnosed ADHD on a similar drawing condition to examine how creativity and knowledge structures are related. Students were asked to construct drawings of alien fruit, and White determined that those with ADHD drew fruit that was based less on known fruit. In the analysis, White concludes that those with ADHD are more flexible in their response to task demands than those without ADHD.

From these experiments, we can see evidence that psychologists already see imagination and flexibility as related. We can conclude that the way in which one organizes information relates to how imaginative they are, and that the better one is able to move information from one setting to another, the more imaginative they are determined to be. This supports supplementing PA-I accounts with the concept of flexibility to explain the processes that lead to the production of an imagination state. As a result, we can make sense of attributions of imaginativeness in reference to this process of being flexible.

Wednesday, 28/Aug/2024 10:20 – 11:00
Philosophy of Language
Future and Probability
Jacek Wawer[1], Jakub Węgrecki[2]
[1]Jagiellonian University, Kraków, Poland; [2]Doctoral School in the Humanities, Jagiellonian University, Kraków, Poland | jacek.wawer@uj.edu.pl

One of the longstanding questions in philosophy is whether propositions about contingent future can be presently true. According to linearists, the answer is 'yes'. They claim that a proposition about the future is true if it is true with respect to the actual future (e.g. Øhrstrøm 2009). On the other hand, universalists claim that the answer is 'no': a proposition about the future is true at a moment m if it is true with respect to all possible futures available at m (e.g. Todd 2021; Thomason 1970).

Quite recently, new arguments emerged which suggest that regardless of which position one endorses, it is going to clash with the intuitive probability assessments. Assume that p is the proposition that it will rain in Prague this week. Intuitively, we should ascribe a non-trivial probability to p (i.e., different from 0 and 1) because it is neither impossible nor settled that it will rain in Prague this week. The puzzle is that neither universalists nor linearists can account for this intuition. Universalists are apparently forced to concede that every future contingent has probability 0 (see Carriani 2021; Todd 2021; Williams 2008), while linearists seem committed to the claim that future contingents have uniquely extreme probabilities - 0 or 1 (see De Florio & Frigerio 2022).

Consider linearists first. If we assume that p is true, then the probability of its being true is equal to 1. However, by T-schema, the probability of p should be equal to the probability of p being true. So, the probability ascribed to p should be equal to 1 (because we assumed that p is true). On the other hand, according to universalists, p is not true. Consequently, the probability of its being true is equal to 0. Again, by T-schema, the probability of p should be equal to the probability of p being true. So, the probability ascribed to p should be equal to 0.

The aim of the paper is twofold. Firstly, will clarify and simplify the puzzles. Secondly, we will provide new solutions to the puzzles and argue that the objections are fatal to neither

universalists nor linearists. We offer two novel reactions available to universalits. One of them is based on Thomason's (1970) supervaluationism, while the other involves a primitive notion of a future-oriented probability, premonitions of which can be found in (Todd 2021). However, our assessment is that the responses available to linearists are less costly than those of their rivals. Our linearists-friendly solution to the puzzle is based on the rejection of the premise that if p is true, then the probability of its being true is equal to 1. The upshot of the argument is that this assumption is unjustified in the light of the linearists' theory because they reject the claim that if something is true, it is settled that it is true. We conclude that the notion of probability fits better with the linearists semantics of future contingents than with their universalist rival.

Selected references

Cariani F. (2021). The Modal Future: A Theory of Future-Directed Thought and Talk. Cambridge University Press.

De Florio, C., & Frigerio, A. (2022). Future, truth, and probability. Inquiry, 1-13.

Øhrstrøm, P. (2009). In Defence of the Thin Red Line: A Case for Ockhamism. Humana Mente 8, 17–32.

Thomason, R. (1970). Indeterminist time and truth-value gaps. Theoria 36(3), 264-281.

Todd, P. (2021). The Open Future: Why Future Contingents Are All False. Oxford: Oxford University Press.

Williams, J. (2008). Aristotelian Indeterminacy and the Open Future. https://philpapers. org/rec/WILAIA4.

Wednesday, 28/Aug/2024 9:40 – 10:20

Philosophy of Language

High hopes for Eternalism

Jakub Węgrecki

Doctoral School in the Humanities, Jagiellonian University, Kraków (Poland) | jakub.wegrecki@gmail.com

Proposition are assumed to be contents of our attitudes. Mark Richard (1981), in his famous paper, argued that temporal propositions, i.e. propositions that change their truth-values over time, are not suitable for this role. He claimed that temporal propositions cannot be the contents of our beliefs. While his argument faced criticism (e.g., Brogaard 2012, Sullivan 2014), the eternalist-temporalist debate remains unresolved.

Recently, new arguments have emerged, shifting the focus from the contents of beliefs to the contents of other attitudes. In this vein, Brogaard (2022) argues that eternal propositions cannot be the contents of future-directed hopes, while Skibra (2021) argues that eternal propositions cannot be the contents of desires. Both of these arguments against eternalism point to issues with semantic satisfaction.

Consider my hope (or desire) that it will be sunny in Prague. According to semantic eternalism, the content of my hope is the proposition that there is a time t later than the time of the formation of my hope such that it is sunny in Prague at t. This proposition is true or false eternally, so it is true or false at the time of my hope's formation. However this implies that my hope is fulfilled (or fail to be fulfilled) at the time of its formation, which seems to be an absurd conclusion. In general, the problem is that given semantic eternalism, our hopes (and desires) are fulfilled (or fail to be fulfilled) as soon as they are formed. In my presentation I will focus on this puzzle. After careful reconstruction, I will show how eternalists can address this objection.

The puzzle rests on two assumptions concerning satisfaction. Firstly, it is assumed that hope (desire) is preferential. Preferential attitudes are satisfied when the world changes to fit the content of the attitude in question. The second assumption (FIT) is that the world changes to fit the proposition p iff there are times t_1, t_2 such that p is false at t_1 and p is true

at t_2. From these two, one can derive the satisfaction condition that generates the puzzle: a hope for p is satisfied at t iff p is true at t.

My solution to the puzzle is to reject the claim that a hope for p (where p is a proposition) is satisfied at t if and only if p is true at t. Instead, we can link the satisfaction of hope for p with truth of a different proposition p´ that is systematically related to p. This strategy mirrors the Sullivan's (2014) temporalist response to Richard's (1981) problem. I will show how this strategy can be adopted by rejecting (FIT). I will propose two alternative accounts of what it means for a world to change to fit a content p. One is based on the B-theoretical account of change (Mozersky 2015). The second account assumes an A-theory picture of time and uses the notion of wide-base truthmaking (Caplan & Sanson 2011: 198). My conclusion is that the case for eternalism is not yet lost.

SELECTED REFERENCES

Caplan, B. & Samson, D. (2011), Presentism and Truthmaking, Philosophy Compass 6(3), 196-208.

Brogaard, B. (2012), Transient truths: An essay in the metaphysics of propositions, New York: Oxford University Press.

Brogaard, B. (2022), Temporal Propositions and Our Attitudes toward the Past and the Future. In C. Tillman & A. Murray (Eds.) The Routledge Handbook of Propositions (pp. 347-362). New York: Routledge.

Mozersky, M. J. (2015). Time, Language, and Ontology: The World From the B-Theoretic Perspective, Oxford: Oxford University Press.

Richard, M. (1981), Temporalism and eternalism, Philosophical Studies 39 (1), 1-13.

Skibra, D. (2021). On Content Uniformity for Beliefs and Desires, Review of Philosophy and Psychology 12, 279-309.

Sullivan, M. (2014), Change We Can Believe In (and Assert), Noˆus 48(3), 474-495.

Monday, 26/Aug/2024 16:30 – 17:10
Philosophy of Mind

A Non-Disjunctivist Account of Reference in Memories from Perceptual and Non-Perceptual Experiences

Markus Werning, Kristina Liefke
Ruhr University Bochum, Germany | markus.werning@rub.de

The paper addresses memories with an episodic character that originate from experiences other than firsthand perceptions. While traditionally philosophers, psychologists, and neuroscientists have concentrated on memories stemming from perceptual experiences, recent attention has been directed towards memories from dreams, hallucinations, vicarious experiences, and fiction (Anonymous, 2023a). Despite differences in their source experiences, psychological and neuro-scientific evidence suggests that all memories with an episodic character share similar underlying causal mechanisms and thus belong to the same natural kind as firsthand perceptual memories.

The paper aims to develop a unified, non-disjunctivist theory of mnemonic reference for both perceptual and non-perceptual episodic memories, utilizing the trace minimalist theory of episodic memory (Anonymous, 2020) and the theory of referential parasitism (Anonymous, 2022). Phenomenological, behavioral and linguistic evidence suggests commonalities between firsthand perceptual memories and memories from non-perceptual experiences, supporting the potential for one natural kind.

Memories from dreams, hallucinations, vicarious experiences, and fiction closely resemble firsthand perceptual memories in their phenomenology. While they may involve different participants and perspectives, such as first or third-person viewpoints, the

differences are not absolute. For instance, some firsthand memories may have third-person perspectives, while dream memories may have first-person perspectives. Behavioral studies comparing vicarious memories to firsthand ones reveal similarities in emotional intensity, physical reaction, vividness, and the sense of "seeing the event" (Pillemer et al., 2015).

Linguistic memory reports also share a number of features – so-called episodicity markers – that can be regarded as indicative of the episodic character of memories (Anonymous, 2023b,a; Umbach et al., 2021). Amongst those linguistic features are adverbial modifiers like *vividly*, *in detail* etc. as well as non-manner *how*-complements Umbach et al. (2021); Anonymous (2023b) and the use of progressive tense. The following examples are reports of perceptual (1), oneric (2), vicarious (3), and fictional memories (4), which highlight the episodic character of the remembered scenarios:

(1) John took a stroll in the park and came across a tall birch tree. A week later, he remembers *vividly how* the tree was sway*ing* in the wind.

(2) Last night, in a dream, Leyla was riding on a horse through a cornfield. When her husband asked her about the dream the next day, she remembered *vividly how* the ears of corn were folding under her.

(3) Anne remembers *empathically how* her uncle was suffer*ing* in a Siberian POW camp. While still alive, he had told her impressively what he went through in the camp. She had listened to his story with great compassion.

(4) While the professor was preparing his seminar on Dostojevski's "Crime and Punishment", a *detailed* memory came up in him of *how* Raskolnikov was nervously hid*ing* in the house of the pawnbroker after he had murdered her and her sister.

We propose a conditional argument: if perceptual and non-perceptual memories with an episodic character indeed share a uniform underlying causal mechanism, then they constitute one natural kind – the natural kind of episodic memories. Consequently, any theory of reference for episodic memories should be non-disjunctivist. However, prevailing theories of reference in the philosophy of mind, such as causalism and descriptivism, are, as we show, inadequate to explain reference in both perceptual and non-perceptual memory cases. Thus, an alternative theory of reference is required to avoid disjunctivism in mnemonic reference.

The issue of reference in episodic memories intertwines with the debate between relationalist and representationalist theories of memory Sant'Anna and Barkasi (2022). Escaping this dichotomy becomes crucial in formulating a comprehensive theory of reference that accommodates both perceptual and non-perceptual episodic memories. Relationalists argue that in remembering, the subject is directly related to the past event or concrete object involved in that event, while representationalists contend that the subject only bears and intentional relation to the past event. This fundamental disagreement leads to differing analyses of the objects of memory and the diachronic nature of episodic memory.

Traditionally, both relationalists and representationalists agree on the diachronic nature of episodic memory, as exemplified by Reid's 2002 Previous Awareness Condition. This condition suggests that an agent can only have an episodic memory of an event if they were aware of it at the time it occurred. Relationalists incorporate diachronicity into their notion of memory inherently, as remembering involves a direct relation to the remembered event. In contrast, representationalists require additional assumptions to account for diachronicity, often invoking preservationist mechanisms to bridge the gap between past experiences and remembering.

Representationalism suggests a *de dicto* analysis of memory reports, where the content of memory is a set of possible worlds containing representations of the past event. In contrast, relationalism advocates for a *de re* analysis, where the content of memory directly involves the actual object or event remembered in the same world.

In conclusion, neither relationalist nor representationalist accounts of episodic memory align with prevalent theories of singular reference such as the causalism or descriptivism. Both *de re* and *de dicto* analyses fail to accurately capture the truth conditions of remembering and misremembering from both perceptual and non-perceptual experiences. A theory of reference is needed that builds upon the reference relation in primary experiences without relying on simple transduction of causal connections or descriptive contents. Referential parasitism, combined with trace minimalism, offers a way forward. This approach provides a uniform, non-disjunctivist, and naturalistic account of remembering and misremembering, regardless of the origin of memory. It also ensures the diachronicity of episodic memory without resorting to problematic preservationist assumptions. By linking remembering to past experiences through minimal, non-representational traces, this framework offers a comprehensive understanding of episodic memory that transcends the limitations of relationalism and representationalism.

References

Anonymous. 2020. – blinded for review – .

Anonymous. 2022. – blinded for review –.

Anonymous. 2023a. – blinded for review –.

Anonymous. 2023b. – blinded for review – .

Pillemer, D.B., K.L. Steiner, K.J. Kuwabara, D.K. Thomsen, and C. Svob. 2015. Vicarious memories. *Consciousness and Cognition* 36: 233–245.

Reid, T. 2002. *Essays on the Intellectual Powers of Man*. University Park, PA: The Pennsylvania State University Press.

Sant'Anna, A. and M. Barkasi. 2022, September. Reviving the naive realist approach to memory. *Philosophy and the Mind Sciences* 3. https://doi.org/ 10.33735/phimisci.2022.9192.

Umbach, C., S. Hinterwimmer, and H. Gust. 2021. German *wie*-complements: Manners, methods and events in progress. *Natural Language and Linguistic Theory* online first. https://doi.org/10.1007/s11049-021-09508-z.

Tuesday, 27/Aug/2024 18:10 – 18:50
Philosophy of Mind

Knowledge-first Dispositionalism about Belief

Simon Wimmer

Technical University Dortmund, Germany | simon.wimmer@hhu.de

I introduce, and highlight the explanatory benefits of, a knowledge-first theory of the folk psychology of belief. I make precise Williamson (2000)'s suggestion that "[...] to believe p is [...] to treat p in ways similar to the ways in which subjects treat propositions which they know" (pp.46-7). The resultant theory says that

    To believe p is
        (i) to treat p in the fully specific way F such that
        (ii) F is relevantly similar to the fully specific way G such that
        (iii) in normal circumstances, anyone treats anything they know in G.

    (i) is a habitual like 'He seems to do 'The Times' crossword' (Huddleston and Pullum, 2002, p.161) and does not require one to treat p in any way at the time at which the belief is held; instead, it ascribes a habit. Thus, my theory is dispositionalist (cf. Ryle, 1949). (iii) then

specifies the dispositional stereotype for belief (cf. Schwitzgebel, 2002) by appeal to knowledge. This appeal to knowledge predicts the developmental trajectory of the folk psychologies of knowledge and belief summarized by Phillips et al. (2020) "Knowledge before Belief". Moreover, since in a knowledge-first framework, knowledge is not reducible to belief, this appeal to knowledge provides an informative characterization of the dispositional stereotype. To arrive at this characterization, we can ask 'In normal circumstances, how does anyone treat everything they know?' Although the complete answer to this question is not straightforward to find, even partial answers, which are straightforward to find (using p as a premise in one's practical reasoning normally accompanies knowing p; so do asserting p, judging p, and not wondering whether p), suffice for my theory to make concrete predictions.

Reference to similarity with a dispositional stereotype predicts that the verb 'believe' admits degree modifiers ('fully/mostly'), comparatives ('more than'), and equatives ('as much as'). Given my theory, such constructions concern the degree to which a subject's fully specific way of treating p resembles the dispositional stereotype. There are two added benefits of my theory's account of the gradability of 'believe'. First, it predicts that, just as Koev (2019, p.5) argues, 'believe' is associated with an absolute standard (cf. 'flat'). Absent contextual manipulation and overt modification, the required degree of similarity to the dispositional stereotype is thus absolute. This explains intuitions to the effect that delusions and superstitions, among others, are not beliefs (Gallagher, 2009; Ichino, 2018). Second, my theory's account of the gradability of 'believe' explains experimental results obtained by Buckwalter, Rose, and Turri (2015). They find that overt modification with 'at least on some level' leads significantly more participants to ascribe belief to protagonists who exhibit hardly any similarities with the dispositional stereotype specified by my theory. This is predicted by my theory: 'at least on some level' relaxes the required degree of similarity.

Reference to relevant similarity in (ii) allows my theory to also predict effects of contextual manipulation. For instance, in a context where interlocutors focus on whether a subject is in the habit of asserting p, this respect of similarity is relevant. Thus, in this context, my theory predicts even a subject that does not use p as a premise in much of their practical reasoning, but does assert p, to believe p. In line with natural case descriptions, patients suffering from the Cotard delusion, who assert that they are dead, are therefore classified as believing that they are, despite their failure to use the proposition that they are dead in much of their practical reasoning (cf. Bortolotti, 2010, p.163). The benefits of this context-sensitivity also generalize to other cases of delusion.

References

Bortolotti, L. (2010). Delusions and other irrational beliefs. International perspectives in philosophy and psychiatry. Oxford: Oxford University Press.

Buckwalter, W., D. Rose, and J. Turri (2015). "Belief Through Thick and Thin". In: Noûs 49.4, pp. 748–775.

Gallagher, S. (2009). "Delusional realities". In: Psychiatry as Cognitive Neuroscience. Ed. by M. Broome and L. Bortolotti. Oxford: Oxford University Press, pp. 245–266.

Huddleston, R. D. and G. K. Pullum (2002). The Cambridge grammar of the English language. Cambridge, UK; New York: Cambridge University Press.

Ichino, A. (2018). "Superstitious Confabulations". In: Topoi.

Koev, T. (2019). "Strong beliefs, weak commitments". In: Proceedings of Sinn und Bedeutung 23.2, pp. 1–18.

Phillips, J. et al. (2020). "Knowledge before Belief". In: Behavioral and Brain Sciences, pp. 1–37.

Ryle, G. (1949). The Concept of Mind. London: Hutchinson & Co.

Schwitzgebel, E. (2002). "A Phenomenal, Dispositional Account of Belief". In: Noûs 36.2, pp. 249–275.

Williamson, T. (2000). Knowledge and Its Limits. Oxford: Oxford University Press.

Naïve Realism and the Mark of the Mental

Paweł Jakub Zięba

Jagiellonian University, Poland | multipawelek@gmail.com

1. According to naïve realism, the phenomenal character of perception is at least partially constituted by the perceived items. Some of the most influential naïve realists claim that this is so because perception is a primitive non-representational relation of conscious acquaintance with mind-independent items in one's environment (Brewer, 2011; Campbell, 2002). Call it 'standard naïve realism' (SNR). While SNR defines perception in relation to consciousness, there's a lot of evidence suggesting that perception can occur unconsciously. But SNR-theorists don't see unconscious perception as a counterexample to their view. They insist that unconscious perception, if it exists, is not an episode of the same fundamental kind as ordinary conscious perception (French & Phillips, 2023; Phillips, 2018).

2. However, the naïve realist can accommodate the claim that episodes of the same fundamental kind as ordinary conscious perception can occur unconsciously. This can be done by specifying naïve realism along the lines of Pure Relationalism (PR), an idea identified by Stoneham (Stoneham, 2008). According to PR, perception isn't a state of the subject, but a relation in the world that enables the subject to form certain beliefs and behave in certain ways. Perceiving is having something before the mind, in the sense that the perceived items are available for reason and judgement, but their being perceived doesn't itself involve any mental act. It follows that perceptual phenomenal character is diaphanous, i.e. entirely constituted by the perceived items (cf. Stoneham, 2008, p. 315). Because the perceived items are mind-independent, and thereby also consciousness-independent, diaphaneity makes it logically possible that perceptual phenomenal character isn't inherently conscious. The PR-theorist can maintain that perceptual phenomenal character is only conscious when the subject becomes conscious of what they are perceptually related to. Therefore, PR predicts that perception can occur unconsciously. Another reason why this alternative formulation of naïve realism is possible is that the main motivations of naïve realism apply to both conscious and unconscious perception (Zięba, 2019).

3. Given the foregoing, one may wonder why any naïve realist would still prefer SNR over PR. After all, specifying naïve realism in terms of PR buys the naïve realist a more unified account of perception. The SNR-theorist may argue that conscious and unconscious perception should be classified differently due to their radically different potentials for action-guidance and knowledge acquisition, but it's unclear whether a difference of this sort is a good reason to consider conscious and unconscious perception as fundamentally different.

4. I will argue that this disagreement has its source in diverging intuitions that SNR-theorists and PR-theorists most likely have about the mark of the mental. Notice first that perception is one of the most obvious examples of the mental. Most philosophers of mind would arguably agree that if the term 'mental' refers to anything at all, it certainly refers to perception. Now, the SNR-theorist construes perception as conscious and non-representational, while resisting the suggestion that an episode of the same fundamental kind as conscious perception can occur unconsciously. This indicates that the SNR-theorist is prone to believe that consciousness is the mark of the mental. According to Stoneham's PR, by contrast, unconscious perception is possible, but perception isn't a mental state (Stoneham, 2008, pp. 310–311, 313). On this view, the mark of the mental can be identified with either consciousness or intentionality (or both), but always at the cost of locating

perception beyond the boundaries of the mental. Because it seems so obvious that perception is a mental phenomenon, the PR-ist conception of perception qua non-mental relation is highly counterintuitive, which is precisely why most naïve realists won't see PR as a genuine alternative to SNR, despite the ability of the former to accommodate unconscious perception.

5. Nevertheless, taking a closer look at various approaches to the problem of the mark of the mental brings one to the conclusion that the best option for the PR-theorist is to embrace the view that there is no mark of the mental. The construal of perception as conscious acquaintance, as well as the ensuing denial to consider unconscious perception as genuine perception, are in perfect agreement with the Cartesian view that consciousness is the mark of the mental. But if one instead construes perception (again, one of the most obvious examples of the mental) as a not essentially conscious non-representational relation (i.e. along the lines of PR), one ends up corroborating the diagnosis of Richard Rorty, who has famously argued that the category of the mental lacks underlying unity (Rorty, 1979).

6. One important consequence of this move is that it completely eradicates the pressure to find a common denominator between all phenomena traditionally regarded as mental. For if Rorty is right that there's no mark of the mental, that pressure is most likely hindering our attempts to explain those phenomena by forcing us to fit them into some unifying explanatory pattern or theory. Most notably, dispensing with the mark of the mental eliminates the pressure to attribute representational content to all phenomena traditionally considered mental (notice that some of those phenomena are less apt to have representational content than others, e.g. moods vs. thoughts). Since one of the main motivations of naïve realism is skepticism about attributing such content to perception, Rorty's approach lends support not only to PR, but also to naïve realism in general.

References

Brewer, B. (2011). Perception and its Objects. Oxford University Press.

Campbell, J. (2002). Reference and Consciousness. Oxford University Press.

French, C., & Phillips, I. (2023). Naïve Realism, the Slightest Philosophy, and the Slightest Science. In B. McLaughlin & J.

Cohen (Eds.), Contemporary Debates in the Philosophy of Mind (pp. 363–383). Wiley-Blackwell.

Phillips, I. (2018). Unconscious Perception Reconsidered. Analytic Philosophy, 59(4), 471–514.

Rorty, R. (1979). Philosophy and the Mirror of Nature. Princeton University Press.

Stoneham, T. (2008). A Neglected Account of Perception. Dialectica, 62(3), 307–322.

Zięba, P. J. (2019). Naïve Realism about Unconscious Perception. Synthese, 196(5), 2045–2073.

FLASH
TALKS

An Alternative Disease Model of Addiction

Saibiayi Alimu
University of Macau, Macau S.A.R. (China) | yc17703@um.edu.mo

1. Introduction

In the realm of psychology and neuroscience, addiction has long been conceptualized as a neurobiological disease characterized by compulsive drug use. In a philosophical viewpoint, it needs a more nuanced characterization of the phenomenon of addiction.

Alan Leshner once proposed a famous account and claimed that addiction is primarily a brain disease resulting from the long-term effects of drug use on brain structure and function (Leshner, 1997). Leshner's account starts the discussion also among philosophers, for instance, Neil Levy responded to this account directly and claimed, "Addiction is not a brain disease (and it matters)." (Levy, 2013)

This article aims to delve into the debate surrounding these viewpoints, examining them through the lenses of the compulsion model of addiction and the choice model of addiction. Additionally, it explores the current addiction models through the framework of theory of mind and suggests adopting an alternative disease model—an externalist perspective on mental illness.

2. Brain disease and non – brain disease debate

Psychologist Alan Leshner, an important figure in addiction research, put forth the disease account of addiction. His main thesis of addiction is as such: addiction is a chronic, relapsing brain disorder characterized by compulsive drug seeking and use. (Leshner, 1997,45)

He supported his thesis with empirical evidence of neuroimaging, highlighting both acute drug use and the long-term drug use alter brain function. This results in distinct difference between addicted brain and nonaddicted brain. Therefore, Leshner argues that addiction is a brain disease.

Meanwhile, Leshner emphasized on the compulsive feature in addiction. He argues that addicts can't control their behaviors just like schizophrenics cannot control their hallucinations and delusions; Parikinson's patients cannot control their trembling.

However, Leshner's viewpoint faces philosophical scrutiny regarding the classification of addiction as a disease solely based on brain differences. His thesis received criticism by Levy in a paper entitled "Addiction is not a brain disease (and it matters)". Levy argues that addiction is a brain disease only under two premises. First, if the neural correlates in addiction are pathological. Second, if the pathology is sufficient for the person to have a disease. (Levy, 2013)

Drawing from philosophical accounts of dysfunction and with the support of neuropsychological evidence, Levy argues that it is true that addiction involves a neuropsychological dysfunction, it does not inherently qualify as a brain disease.

Levy also questions whether addiction sufficiently impairs an individual's agency functioning. He argues that the former heavy drinker or drug taker who has been abstinent for many years need not be suffering from any impairment, and most people addicted to caffeine do not exhibit significant impairment.

If Levy's argument holds true --- that the neural correlates of addiction do not necessarily lead to impairment, then addiction cannot be classified as a brain disease in the manner proposed by Leshner. Unlike conditions like schizophrenia and Alzheimer's disease, which entail profound defects in rationality and agency, addiction, according to Levy, may involve deficits in self-control mechanisms but does not necessarily result in significant impairments of agency or rationality.

3. Alternative disease model

The controversy between Leshner and Levy stems from their adherence to different models of addiction, the compulsion model and the choice model.

The compulsion model posits that addicts are not capable of freely acting upon their behaviors due to compulsion. Leshner is a prominent advocate of this perspective. Proponents of the choice model view addiction as a voluntary behavior and an act of choice.

Levy contends that addiction arises when individuals lack the cognitive resources to resist temptations, suggesting that addicts still exercise control over their actions based on their judgments. (Levy, 2014)

This conflict between Leshner and Levy is rooted in differing metaphysical perspectives—reductionistic versus dualistic. If our mind is identical and entirely determined by the neural processes in our brain, then compulsion in addiction is reducible to changes in brain function. If mind is a separate entity from the body as dualists would argue, then people still act freely in their action. Addicts can make choices based on their judgments.

In this regard, this debate over whether addiction is a brain disease encompasses two distinct aspects:: firstly, whether addiction is a disease of the *brain* , meaning whether it can be fundamentally characterized as a neurochemical condition. Secondly, whether it fits the classification as a *disease* .

Levy's counterargument seems to conflate the notions of disease and illness, particularly in a dualistic perspective where brain disease does not necessarily equate to mental illness. He argues that addiction is not a brain disease do not mean that addiction is not a disease at all. To be more specifically, he cannot reject that addiction is a mental illness.

An externalist model in psychology and psychiatry proposes that a person's mental health depends on social and cultural factors, not solely on brain-related activity.

Addicts may experience severe anxiety and depression, leading them to self-medicate with substances (Khantzian, 1997) Moreover, many addicts endure socio-economic challenges and poor mental health (Pickard, 2019) These findings advocate for adopting an externalist perspective rather than the traditional dichotomy of brain disease versus non-brain disease models of addiction.

4. Conclusion

I began by presenting the debate about whether addiction is a brain disease, as articulated by psychologist Alan Leshner and philosopher Neil Levy. Leshner's proposition aligns with the compulsion model of addiction, categorizing it as a brain disease. Levy's perspective falls under the choice model, suggesting addiction involves voluntary behavior. The crux of these models lies in the question of whether addicts possess free control over their actions. Moreover, their disagreement stems from opposing metaphysical viewpoints: reductionistic versus dualistic.

Despite disagreeing with the idea that addiction is solely a brain disease, an alternative perspective may consider addiction as a mental illness without entirely reducing it to brain circuitry. I argue that the externalist perspective can be an alternative disease model of addiction, which emphasizes the role of social, cultural, and environmental factors alongside neurological processes in shaping addictive behaviors.

Leshner Alan. (1997) Addiction is a brain disease, and it matters. Science. Oct 3;278

Levy N. (2013) Addiction is Not a Brain Disease (and it Matters). Front Psychiatry. Apr 11

Levy, Neil. (2014) Addiction and Compulsion in O'Conner, Timothy ; Sandis Constantine (Eds) A companion to philosophy of action, A John Wiley & Sons, Ltd., Publication, 267-273

Pickard, Hanna. (2018). THE PUZZLE OF ADDICTION. In Pickard, Ahmed (Eds) Routledge Handbook of Philosophy and Science of Addiction Routledge

Khantzian, E. J. (1997) "The self-medication hypothesis of substance use disorders: a reconsideration and recent application", Harvard Review of Psychiatry 4(5): 231–244.

Philosophy of Language

"Desolate" as a Case Study in Non-Canonical Descriptive Slurs

Michaela Fikejzová

Faculty of Humanities, Charles University, Czech Republic | michaela.fikejzova@gmail.com

"Just because he doesn't have teeth, doesn't mean he's a desolate, just because he's an asshole without an opinion usually does," states an anonymous profile on the social network *X* during a debate regarding Ukrainian refugees. The expression "desolate" began to appear in Czech media and public discourse in 2022, specifically in relation to the anti-government demonstrations that took place in response to radically increasing inflation and rising energy prices. These demonstrations carried strong anti-EU and anti-NATO undertones, contributing significantly to the definition of the individuals participating in the demonstrations. These persons began to be referred to as "desolates," a term with no stable semantic dictionary definition. *The Academic Dictionary of Contemporary Czech* (2024) defines a "desolate" as "a person who is unsuccessful in life because of his or her bad character traits, a loser (often derogatory)." In contrast, the community-managed dictionary of neologisms *Czech 2.0* (2024) defines a desolate person as "(i) a homeless person or a drunkard, very repulsive in appearance, smell, etc.; (ii) a protagonist of the disinformation scene, usually a permanently dissatisfied individual on the fringes of society, who compensates for his frustration in life by rejecting social authority and turning to conspiracy theories and political extremism; (iii) a derogatory term." *Wiktionary*'s (2024) first definition gives a similar characteristic to (ii) in the case of *Czech 2.0*: *"(neologism, pejorative)* a person who spreads and falls for disinformation; usually rejects the authority of the state and state institutions, is prone to conspiratorial thinking" and the second definition is the *Academic Dictionary of Contemporary Czech*'s definition. A common intersection of dictionary definitions is the designation of the term as derogatory or pejorative. The present paper aims to show whether the newly arisen derogative or pejorative term "desolate" can be considered a slur. In the expressivist framework proposed by Robin Jeshion (esp. 2013, 2017, 2021), I will argue for this expression being a pejorative lexical item, using primarily the most extreme form of their use, i.e. their use as a pejorative speech act performed with a pejorative lexical item used pejoratively (cf. Jeshion 2021: 213). Specifically, I consider it a slur based on stereotyping other than gendered, racial, and ethnic stereotyping. Stereotyping related to the notion of the "desolate" is based on an axis of presumed intelligence derived from specific political beliefs intersecting with the axis of economic class. From a philosophy of language perspective, the notion of "desolate" is a specifically borderline example not only because of its novelty and relative lack of an accepted definition but also because it violates the principle of neutral counterparts questioned by Foster (2023). Later in the paper, therefore, I develop Foster's argument while using an example (seemingly) completely devoid of its semantic neutral counterpart. The final part of the argument is to relate the notion of "desolate" to the act of reclaiming a slur, which in media and public discourse runs parallel to the granting or denying of slur status to the term itself. As I will argue, it is a pride reclamation in Jeshion's (2020) sense. The philosophical analysis is based on a sample of media outputs from the nationwide Czech

media between March 2022 and December 2023, which explicitly seek to establish a definition of the pejorative and discuss the pejorative aspects of the label or their absence. In addition, a sample of posts on social network *X* from the period September 2023 - December 2023 that actively use the term "desolate" either as a non-canonical stereotyping slur, i.e. pejoratively, or are an active act of pride reclamation is used.

Foster, J. (2021): Busting the Ghost of Neutral Counterparts, *Ergo. An Open Access Journal of Philosophy*, 10(42): 1187-1242.

Jeshion, R. (2013): Expressivism and the Offensiveness of Slurs, *Philosophical Perspectives*, 27(1): 231-259.

Jeshion, R. (2017): Slur Creation, Bigotry Formation. The Power of Expressivism, *Phenomenology and Mind*, 11: 130-139.

Jeshion, R. (2020): Pride and Prejudiced. On the Reclamation of Slurs, *Grazer Philosophische Studien*, 97: 106-137.

Jeshion, R. (2021): Varieties of Pejoratives, in Khoo, J, & R. K. Sterken: *The Routledge Handbook of Social and Political Philosophy of Language*, 211-231.

https://cestina20.cz/slovnik/dezolat/

https://slovnikcestiny.cz/heslo/dezolát/1/49367

https://cs.wiktionary.org/wiki/dezolát

Philosophy of Language
Explaining: A Normative Account
Grzegorz Gaszczyk[1], J.P. Grodniewicz[2]
[1]University of Warsaw, Faculty of Philosophy, Poland; [2]Jagiellonian University, Copernicus Center for Interdisciplinary Studies, Poland | gaszczyk.grzegorz@gmail.com

From the origins of black holes to the flow of electorate between presidential candidates and the rise in popularity of K-pop — the world is complicated. And yet, we do our best to understand it. In some cases, each of us tries to do it on our own. But with the majority of phenomena, we either need to rely on others willing and able to explain things to us, or we are the ones offering explanations. Whenever we do that, we take part in the practice of explaining.

We argue that, while the practice of explaining is unified by its function — dissemination of understanding (Lipton 2004; Grimm 2010) — there are different ways in which we can engage in it. We can do it by offering what we call a minimal explanation, a default explanation, or a full-blown explanation. Each of them is appropriate in different contexts and comes with a set of relevant expectations. Moreover, each of them is guided by a specific norm.

We start by characterizing what we call the minimal explanation. A natural initial assumption is that to explain something, one needs — minimally — to understand it (Achinstein 1983). This is reflected in the following norm, which we call UNDERSTAND, or (U) for short:

(U) One should explain p only if one understands p.

We believe that (U) demarcates an entry point to the practice of explaining. Even though we consider it rather uncontroversial, we briefly discuss two cases that may challenge (U) as the norm of minimal explanation.

The first one is the case of a substitute teacher, who teaches evolution despite not understanding it, simply by reading a chapter from a biology book. We argue that the teacher does not explain — they serve merely as a proxy transmitting the explanation offered by the author of the book (cf. Ludwig 2020).

The second case is one in which someone explains something that they barely understand. They have some of the facts right, and they grasp some connections between them, but they are also wrong with a lot of things. In this case — we argue — the explainer satisfies (U). This is because understanding is gradable (Kvanvig 2003; Elgin 2007), and it would be unrealistic to demand that one should explain p only if one fully understands p — if that were the case, no one would, for example, be in a position to explain the physics of black holes.

Fulfilling (U) is enough to participate in the practice of explaining. However, simply understanding the subject matter is insufficient in most contexts. Think about a university professor invited to a local high school to lecture on the formation of crystals. She has an excellent and in-depth understanding of this phenomenon, in fact, she is one of the leading scholars in the field. Unfortunately, she is not able to adjust the explanation to her audience. Most students are already lost at the third slide, and nobody manages to follow the whole lecture.

This case illustrates that we are typically expected to adjust or customize the explanation to our audience. Hence, we propose the following UNDERSTAND&CUSTOMIZE (U&C) norm of default explanation:

(U&C) One should explain p only if (i) one understands p, and (ii) the means one uses are customized to the audience.

Notice that the customization of the means would not always equal making the material easier and more accessible (cf. discussion on the scientific and expert testimony, e.g., De Regt 2017; Gerken 2022). If the same university professor presents her work to other scientists, she may fulfill the norm by assuming that her audience possesses a great amount of background knowledge and skipping over most of the material to get straight to her novel point.

Many cases of online explanations, such as YouTube videos, pre-recorded lectures, educational podcasts, etc., are crafted for the intended audience (e.g., interested and motivated non-experts) and thus fulfill (U&C). Even though, in those cases, there is no interaction between the explainer and the recipient of the explanation, default explanations often succeed in disseminating understanding.

In many contexts, however, we expect the explainer to be additionally attentive to the audience in real-time. Think about a competent teacher who understands evolution and explains it to her students in a way customized to the level of their background knowledge but who does not monitor whether the student's understanding increases or not. Clearly, despite fulfilling (U&C), the teacher would fail to participate in the practice of explaining in the way expected given their profession. What we expect good teachers to do is to offer what we call the full-blown explanation, governed by the norm UNDERSTAND&CUSTOMIZE&MONITOR (U&C&M).

(U&C&M) One should explain p only if (i) one understands p, (ii) the means one uses are customized to the audience, and (iii) one monitors the audience's understanding of p.

Monitoring is pivotal for tracking the increase of the audience's understanding and identifying what information would be most helpful to introduce at any given moment. We will show that monitoring often happens naturally — explainers do it automatically (e.g., Vélez et al. 2023), and is widespread — there is a variety of educational techniques and theories, like the scaffolding theory, that are intended to ensure successful monitoring (e.g., Reiser & Tabak 2014).

Only if we take all these ways of participating in the practice of explaining into account will we be able to spell out what is expected from explainers in different situations. As a result, we will be able to tell good explainers from bad ones by assessing whether they fulfill

the norm and thus offer explanations (either minimal, default, or full-blown) appropriate for a given context.

References

Achinstein, P. (1983). The Nature of Explanation. OUP.

De Regt, H. W. (2017). Understanding Scientific Understanding. OUP.

Elgin, C. Z. (2007). Understanding and the Facts. Philosophical Studies, 132(1), 33–42.

Gerken, M., (2022). Scientific Testimony: Its roles in science and society. OUP.

Grimm, S. R. (2010). The Goal of Explanation. Studies in History and Philosophy of Science Part A, 41(4), 337–344.

Kvanvig, J. L. (2003). The Value of Knowledge and the Pursuit of Understanding. CUP.

Lipton, P. (2004). Inference to the Best Explanation. Routledge.

Ludwig, K. (2020). Proxy Assertion. In S. Goldberg (Ed.), The Oxford Handbook of Assertion (pp. 305–326). OUP.

Reiser, B. J., & Tabak, I. (2014). Scaffolding. In R. K. Sawyer (Ed.), The Cambridge Handbook of the Learning Sciences (pp. 44–62). CUP.

Vélez, N. et al. (2023). Teachers recruit mentalizing regions to represent learners' beliefs. PNAS, 120(22).

Philosophy of Language
Structures of Inner Discourse: A Strategy Based Approach
Mathijs Geurts
University of Salzburg, Austria | mathijs.geurts@gmail.com

People can do a lot of things by talking. They can use speech to narrate, deliberate, plan, reason or reflect. People can also do a lot of things by talking to themselves. They can use inner speech to narrate, deliberate, plan, reason or reflect. Putting it this way, there seems to be a symmetry between some social and psychological processen. An explanation for this, first formulated by the psychologist Lev Vygotksy (1986), is that higher-order thought results from the internalization of social interactions. In recent years, internalization has been used to explain various cognitive processes, including meta-cognition, reasoning and planning (Frankish, 2018; Kompa & Mueller, 2022; Gauker, 2018). However, what exactly is internalized has not been spelled out in detail.

In the literature on inner speech, it is usually held that what is internalized is a kind of dialogical structure. One proposal for such a structure is presented by Fernyhough (1996). On his account, fully internalized inner dialogue involves flexible coordination of multiple perspectives within an individual. This coordination need not involve the same semantic and syntactic elements as overt speech and as a result, inner speech can appear to be unlike out-loud conversation, though underlying cognitive processes remain dialogical. A problem with this account is that it is unclear how these different perspectives are represented, and how they are coordinated if not by means of conducting a fully worded inner conversation.

Kompa & Mueller (2022) present an alternative account. They interpret inner dialogue as a looping structure that emerges through engaging both speech production and comprehension. This structure is most effectively engaged when an individual adheres to pragmatic principles that govern interpersonal dialogue and is useful for deliberative thinking and reasoning. Contrary to Fernyhough, this account holds that a kind of 'to and fro' is essential for these functions of inner speech. To the extent that any thinking process engages both speech production and comprehension mechanisms, it may be said to be dialogical. The problem with this is that speech production and comprehension are presumably also engaged in producing a complex monologue, which means that there is nothing specifically dialogical about the mechanism they propose.

In addition to these problems, the notion of internalized dialogue may not be saying very much to start with. Dialogue is a broad notion that includes many types of discourse and its open-ended structure may not be sufficient to explain specific types of higher-order thinking.

In order to get a better grip on the types of structure involved in inner speech, I borrow some ideas from the literature on discourse structure and planning. With these ideas, I argue, it becomes possible to describe some functions of inner speech in more detail. This is important because at a basic level, dialogue does not have much structure apart from some form of to and fro. Moreover, this structure can often be removed without much effect. Take the following example:

A1: Who won?
B1: Bob won, but he wasn't happy about it.
A2: Why not?
B2: He thought he should have won by a bigger margin.
A3: What did he do afterwards?
B3: He went into the dressing room and sulked for about half an hour.

Here, A1, A2 and A3 can be removed, turning it into a monologue without loss of information or coherence. This is because the structure of any particular interaction is determined by the strategies employed by interlocutors to achieve individual or collective goals (Carlson, 1983), and those strategies may or may not be represented dialogically. Therefore, strictly speaking it is goal-directed strategies, rather than dialogical structures per se, that explain the higher-order functions of inner speech. In 1), the goal is to establish a general narrative of what happened, and this may include a broad range of information. As a result, what is relevant is that Bob won, but also the contrastive information that he was unhappy about it, the reason for his unhappiness and the sequence of events that followed. Note also that the second and third question-answer pairs can be swapped round without loss of coherence. Compare this with the following adapted from Grosz (1981):

E0: First you have to remove the flywheel.
F1: How do I remove the flywheel?
E1: First, loosen the two allen-head set screws holding it to the shaft, then pull it off.
F2: OK... I can only find one. Where's the other one?
E2: On the hub of the flywheel.
F3: That's the one I found. Where's the other one?
E3: About ninety degrees around the hub from the first one.

In 2) swapping round any of the question-answer pairs would render the sequence incoherent. This is because the interaction is driven by a plan that includes strictly ordered goals and subgoals, which results in a more hierarchical structure. Because of these differences in structure, I argue that an explanation of inner speech function should be based on the goal-directed strategies employed, rather than on general aspects of communication like those referenced by Fernyhough (1996) and Kompa & Mueller (2022). I illustrate this by discussing several real-life examples of private speech being used for action planning, reasoning and narration, and showing how their structure is determined by goals and strategies.

References

-Carlson, L. (1983). *Dialogue Games*. D. Reidel Publishing Company.

-Fernyhough, C. (1996). The dialogic mind: A dialogic approach to the higher mental functions. *New Ideas in Psychology*, 14(1), 47–62.

-Frankish, K. (2018). Inner speech and outer thought. In Langland-Hassan, P., & Vicente, A. (eds.). *Inner speech : new voices.* (pp. 221-243). Oxford University Press.

-Gauker, C. (2018). Inner speech as the internalization of outer speech. In Langland-Hassan, P., & Vicente, A. (eds.). *Inner speech: new voices.* (pp. 53-77) . Oxford University Press.

-Grosz, B. (1981). Focussing and description in natural language dialogues. In Joshi, A. Webber, B. & Sag, I. (eds.) (pp. 84-105). *Elements of Discourse Understanding*. Cambridge University Press.

-Kompa, N., & Mueller, J. (2022). Inner speech as a cognitive tool—or what is the point of talking to oneself? *Philosophical Psychology*, 1–24.

Philosophy of Mind

Mapping Out the 4E Cognition Spectrum
Henrique Mendes Gonçalves

Central European University, Austria | goncalves_henrique@phd.ceu.edu

Over the past three decades, the ideas of cognition as an embodied, embedded, extended, or enactive phenomenon have been widely discussed among philosophers and cognitive scientists. Together, these are believed to constitute a family of very closely related approaches that has been called "4E cognition." (Newen, De Bruin & Gallagher 2018) The expression seems to tacitly introduce three presuppositions:

(1) First, that each "E" stands for a well-defined, unique theory or field of research.

(2) Second, that compatibility issues between these different approaches are a matter of detail, and that the adoption of one of them usually implies the acceptance of the others.

(3) Third, that E-approaches share a set of commitments which is at odds with traditional theories in cognitive science and philosophy of mind, namely: i) antirepresentationalism and the rejection of ii) functionalism, iii) computationalism, and iv) multiple realizability.

Upon closer inspection, however, none of these seems to hold. In this presentation, I will investigate the internal theoretical diversity within each one of the four Es, as well as their external points of incompatibility with each other. My arguments will conform to the following outline:

(1) E-approaches usually present *stronger* and *weaker* versions, which can be individuated based on their stance on whether body and environment play a *constitutive* or merely *causal* role in cognitive processes. This strategy is inspired by Gallagher's account, but I depart from him in applying the weaker-stronger distinction to the fields of embedded and extended cognition too, which I conceive as two varieties of the same theory. (Gallagher 2017, 2023) Thus, in the case of embodied cognition, the weaker variant will incorporate the body into explanations of cognitive and perceptual processes only insofar as it plays a causal role in these processes; the stronger, on the other hand, will introduce brain-body couplings as an ineliminable part of such explanations. Similarly, embedded and extended cognition, I will argue, should be seen as the weaker and stronger variants of the same theory: in the first case, brain-body-environment couplings may play a relevant role in cognitive, perceptual, and motor processes, but only a causal one; in the second, they may also be of constitutive relevance. As we will see, however, these distinctions cannot be applied to the field of enactive cognition, precisely because brain-body-environment couplings are always taken to be constitutive in such theories, which makes weak enactivism an unstable theoretical position.

(2) Additionally, I will investigate the external relations that hold among each of the Es in the 4E spectrum, showing that the weaker-stronger distinction can be used to assess

their mutual compatibility. I will show that the compatibility scheme for 4E cognition theories follow a hierarchical structure: generally, weak extension theories (i.e., embedded cognition) are compatible with both weak and strong embodiment theories; strong extension theories (i.e., extended cognition proper) are compatible only with strong embodied cognition. Again, since it is bereft of a weaker version, enactivism tends to be generally incompatible with weaker theories in the 4E cognition spectrum.

(3) Finally, I will show that some influential E-theories are still very much committed to at least some of the tenets in i-iv)—i.e., i) representationalism, ii) functionalism, iii) computationalism, and iv) multiple realizability. Usually, weaker versions—i.e., those without the constitution claim—tend to be more easily incorporated into such classical cognitivist frameworks. I will illustrate my argument by briefly discussing, in the case of embodied cognition, Goldman & Vignemont's concept of B-formatted representations, and compare it with conceptions of bodily-based representation in cognitive linguistics and in Barsalou's perceptual symbols system. (Goldman & Vignemont 2009; Goldman 2014; Lakoff & Johnson 1999; Barsalou 1999) A similar argument will then be developed for the case of weak and strong extended cognition: embedment theories such as Hutchins' distributed cognition and Gibson's ecological psychology will be compared to extension theories, particularly Chalmers & Clark and Clark's externalist interpretation of results in predictive processing. (Hutchins 1995; Gibson 1979; Chalmers & Clark 1998; Clark 2016) We will also see, however, that both main varieties of current enactive theories—i.e., Noë & O'Regan's sensorimotor enactivism and Thompson's autopoietic enactivism—do indeed reject i-iv). (O'Regan & Noë 2001; Noë 2004; Thompson 2007)

After showing that incompatibility with classic cognitivism and computationalism more broadly construed cannot be wholesale applied to the entire field of 4E cognition, my conclusion will be that features of enactivism, the most radical theoretical family in the 4E cognition spectrum, are often mistakenly assigned to the whole 4E cognition spectrum. In general, embodied, embedded (weakly extended), and extended (strongly extended) cognition theories present various degrees of compatibility with classic cognitivism, accepting or rejecting different elements in i-iv)—but almost never all of them at once. Thus, only enactivism, which rejects i-iv) entirely in all of its current theoretical strands, should be strictly speaking conceived of as promoting a significant paradigm shift in the cognitive sciences and philosophy of mind.

Bibliography:

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.

Chalmers, D.; & Clark, A. (1998). The extended mind. *Analysis*, 58(1), 7–19.

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press.

Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford: Oxford University Press.

Gallagher, S. (2023). *Embodied and enactive approaches to cognition*. Cambridge: Cambridge University Press.

Gibson, J. (1979). *The ecological approach to visual perception*. London: Psychology Press.

Goldman, A. (2014). The bodily formats approach to embodied cognition. In: U. Kriegel (ed.), *Current controversies in philosophy of mind*. New York and London: Routledge, pp. 91–108.

Goldman, A.; & Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences*, 13(4), 154–159.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Lakoff, G.; & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.

Newen, A.; De Bruin, L.; & Gallagher, S. (eds.). (2018). *Oxford handbook of 4E-Cognition*. Oxford: Oxford University Press.

Noë, A. (2004). *Action in perception*. Cambridge, MA: MIT Press.

O'Regan, J.; & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(5), 939–973.

Thompson, E. (2007) *Mind in life: biology, phenomenology, and the sciences of mind*. Cambridge, MA: Harvard University Press.

Philosophy of Language

What Distinguishes Online and Offline Hate Speech? A Transformativist View on Online Abuse

Tomáš Koblížek

Institute of Philosophy, Czech Academy of Sciences, Prague, Czech Republic | koblizek@flu.cas.cz

Recently, the question of how hate speech differs in offline and online spaces has been raised in applied philosophy of language. One of the main contributions to the debate has been provided by Michael Barnes (2023), who has distinguished between two types of hate speech.

Barnes speaks of: 1. Collective subordinating speech acts, characterized by crowd anonymity, which gain authority as a result of the active efforts and mutual solidarity of the hate speech authors (mutual quotes, likes, hastag sharing, language imitation). 2. Individual subordinating speech acts that gain authority passively, through accommodation, i.e. only by not being blocked or opposed to.

The first type, according to Barnes, characterizes *typically* online hate speech: an example is a hate campaign on the X network, where verbal attacks on LGBT people accumulate into a single linguistic entity causing a specific harm. The second type characterizes *typically* offline hate speech: an example is an offensive statement about LGBT people made by a politician during an election rally.

Barnes' distinction is an important step towards understanding online hate speech insofar as it shows the role of the collective: as Barnes accurately points out, hate speech in the online space is not only individual speech, but thanks to the technical aspects of social media, it can also produce specific speech acts of a group nature.

In my article, however, I will show that this conception must be substantially supplemented or modified. More specifically, I will argue that Barnes' conception can be described as *additivist*: on the additivist view, a collective subordinating act is added to individual subordinating acts as a new linguistic act. This new act is qualitatively different from individual speech acts and does not transform their quality.

In contrast, I will defend a conception that I will label *transformativist*: on the transformativist view, when a collective subordinating act is constituted, not only a new linguistic act appears, but the whole linguistic environment is transformed, i.e. individual subordinating acts retroactively acquire new properties. I will focus on the two main properties thus acquired:

1. Individual subordinating act is not only part of, but also a concrete representation, avatar or embodiment of a collective subordinating act. The fact that individual acts are representations, not just parts of a subordinating act, intensifies their subordinating power.

2. Once a collective subordinating act is established, the individual subordinating acts do not have the same authority. Some of them take on more authority insofar as they more significantly determine the dynamics of the entire collective subordinating act. This concerns, for example, the individual acts that introduce a higher level of aggression or a new type of offensive vocabulary further accepted in the collective.

My argument for the transformativist conception of online hate speech rests on the assumption that any account of this phenomenon should be able to explain for four facts associated with this type of speech. I refer to these phenomena as (i) plurality of effective means, (ii) rational response of recipients, (iii) double harm, (iv) and the dynamics of online hate speech.

The plurality of effective means. The authority of online hate speech is routinely undermined both by means of restricting collective or group aspects of hate speech (limiting the possibility of sharing, blocking hashtags, restricting certain types of vocabulary) and by means of targeting individual speech (countering the prejudices or misinformation supporting hate speech, drawing attention to the consequences of hate speech, etc.).

The rational response of the recipients. Despite the collective nature of hate attacks in the online environment, we often consider the tendency of the attacked to respond to individual hate speech to be rational.

Double harm. In online hate speech, there are harms caused by the collective subordinating act, but also harms arising from individual speech.

The dynamics of online hate speech. When someone is continuously attacked in a hate campaign, not only is the collective authority of individual discourses intensified, but also the vocabulary and character of the hateful illocutionary acts are often continuously transformed.

I will show how the transformativist model of online hate speech can account for these phenomena and why the additivist model can do this only imperfectly.

Philosophy of Mind

Emotional Behavior and Contagious Imitation: Cases Between Reflexes and Actions
Zixuan Liu

Husserl-Archiv der Universität zu Köln, Germany | liuzixuan97@126.com

In current action theory, the oversimplified dichotomy of human behavior – represented by various dual-process models which divide human behavior into a 'controlled, flexible, effortful, reflective and conscious' system and a 'automatic, inflexible, effortless, reflexive and unconscious' one (Gawronski & Creighton 2013) – is increasingly under criticism (Salin & Phillips 2007; Fridland 2014, 2017a, b). "The facts that one could unify in the brace of the 'mechanical' are not so homogeneous as one believes"; rather, "the automatism is not a so simple phenomenon as it appears" (Ricoeur 2009, Vol. 1, p. 381, p. 383).

The more appropriate characterization of human behavior is a *spectrum*, with reflexes and 'full-blooded actions' which involve reflective control such as deliberation and decision-making (Velleman 1996) on two extremes. There are a number of in-between cases that are irreducible to each extreme. One of them is habitual action, such as drinking a cup of coffee every morning 'purely out of habit' (Pollard 2006; Douskos 2017). Arguably, this is still an action, insofar as it takes place by virtue of agentive power exertion for certain reasons, but reflective self-determination is not necessary in this case.

Another case is what I call 'expressive behavior', which includes non-purposeful emotional behavior – such as laughing out of joy, crying out of sorrow, stamping one's feet with anger – and contagious imitation: one yawns when he sees another person yawning (Jeannerod 2006, pp. 121–124). Such behavior is less than action, because it is not goal-directed. But it is more than mere reflex, so long as the latter occurs before one is conscious of the trigger, which is not the case in expressive behavior.

Where is expressive behavior positioned in the spectrum of human behavior? If the impression above qualifies as evidence against its reducibility to reflex or habitual action, how should we articulate *the boundaries between reflex, expressive behavior and habitual action*?

Two parameters appear to be promising candidates for this demarcation. One is *the different roles of consciousness* in three cases. Phenomenal consciousness is epiphenomenal in reflexes (Ricoeur 2009, Vol. 1, p. 293). It only serves the post hoc comprehension when one withdraws his hand reflexively from the fire. By contrast, as I would argue against epiphenomenalism regarding actions, factors that are guiding the agentive power exertions, in order to be reasons for action, have to be conscious. Here, goals are always part of the reason. In comparison, expressive behavior does seem to be 'mediated' by consciousness, but the goal is not one of the guiding factors.

The other candidate is *causation*. Reflexes obey the law of event causation, insofar as it can be described in terms of regular successions of events. By contrast, according to the agent-causalist tradition, the relevant causation for action performance is irreducible to event causation. It is intriguing, then, to consider a unique type of causation for expressive behavior.

My proposal means to offer a more precise characterization of these two parameters:

Expressive behavior non-purposefully externalizes the emotional subject's concern in a quasi-symmetrical way.

As for the emotional subject, 'self' is a too 'thick' notion to be simple and primitive (Textor 2018). Rather, it is grounded by more fundamental notions of subjects. One of them is what I call the emotional subject. It is the 'heart' upon which the valuable entities 'lie' with their weights (Ricoeur 2009, Vol. 2, pp. 125–135), considering the German expression for 'concern' (*Anliegen*) which literally means 'lying on one's heart' and the Greek root of value (ἄχιος) which is originally connected with 'weight'.

I use 'concern' to designate the relation between the emotional subject and what concerns it. 'Concern' means that the emotional subject is 'charged' (*laden*) with the load of value or 'impressed' by it, and expressive behavior is what 'discharges' (*entladen*) or 'expresses' this load. But this discharge or externalization is not arbitrary. Rather, the relation between the expressive behavior and its performer is quasi-symmetrical to the one between the emotional subject and what concerns it. Acute pain is expressed by abrupt alteration of behavioral patterns. Those at a loss behave diffusely and randomly. Appall results in fossilization and a collapsed person looks 'fallen apart'. Expressive behavior also concerns the emotional subject and thereby provides either positive or negative feedback to the axiological load. That is a fact captured by the James-Lange theory of emotion.

This proposal addresses the distinctive role of consciousness and the unique sort of causation in expressive behavior. The consciousness is not epiphenomenal in expressive behavior, insofar as how one is conscious of the valuable entities specifies the manner of performance. Unlike in the case of action, consciousness plays no role of justification. For this reason, the relevant causation for expressive behavior is not agent causation, but it is irreducible to event causation as well, which lacks such consciousness-mediated quasi-symmetry.

An important implication: the consciousness-mediated quasi-symmetry lays the foundation for intersubjective understanding in general, because it establishes the original association of one's mental state and bodily behavior and provides the readiness to be impressed by the expressive behavior of others in the case of contagious imitation. This is the prerequisite for sophisticated intersubjective understandings such as communicative actions.

Douskos, C. (2017). Pollard on Habits of Action. *International Journal of Philosophical Studies* 25(4), 504–524.

Fridland, E. (2014) They've lost control: reflections on skill. *Synthese* 191, 2729–2750.

Fridland, E. (2017a). Skill and motor control: intelligence all the way down. *Philosophical Studies* 174:1539–1560.

Fridland, E. (2017b). Automatically minded. *Synthese* 194, 4337–4363.

Gawronski, B. & Creighton, L. A. (2013). Dual Process Theories. In D. Carlston (ed.) *The Oxford Handbook of Social Cognition*.

Jeannerod, M. (2006). *Motor Cognition: What Actions Tell the Self*. Oxford: Oxford University Press.

Pollard, B. (2006). Explaining Actions with Habits. *American Philosophical Quarterly* 43(1), 57–69.

Ricoeur, P. (2009). *Philosophie de la volonté*. Paris: Édition Points.

Saling, L. L. & Phillips, J. G. (2007). Automatic behaviour: Efficient not mindless. *Brain Research Bulletin* 73(1–3), 1–20.

Textor, M. (2018). Enjoy yourself: Lotze on Self-consciousness and Self-Concern. *History of Philosophy Quarterly* 35(2), 157–179.

Philosophy of Language

Demonstratives and the Semantic Redundancy of Intentions

Jakub Rudnicki

University of Warsaw, Poland | jmrudnicki@gmail.com

Intentionalism about demonstratives, such as "that," is the view in which the reference of these expressions' uses is determined, at least partially, by the speakers' referential intentions. This view manifests in two main forms: semantic and pragmatic. The semantic version argues that demonstrative reference is a semantic phenomenon achieved by pairing the expressions' linguistic meaning (for example, Kaplanian characters) with their context of use (King 2014, Stokke 2010, Radulescu 2019). The pragmatic version, conversely, posits that the meaning of demonstratives does not determine their reference based on context. Instead, it argues that it is not the expressions in context that refer, but rather the speakers themselves (Bach 2017, Heck 2014, Smit 2012).

Proponents of the pragmatic approach argue against the semantic version by highlighting the alleged redundancy of its crucial aspect (Heck 2017: 344, Smit 2012: 52). They contend that it is illogical for the semantic convention governing the uses of demonstratives to require hearers to interpret the speakers' intended referents, as this requirement is already inherent in linguistic communication. Therefore, they conclude that the linguistic meaning of demonstratives underdetermines their reference, and that the gap is filled by general pragmatic reasoning.

In my presentation, I will focus on two main objectives. First, I will present a straightforward argument demonstrating why the aforementioned reasoning, despite its intuitive appeal, is fallacious. This argument cannot be universally valid because, if it were, it would also apply to complex denoting expressions like "the object I intend to refer to." However, considering that this phrase clearly refers to the object its speaker intends, as a matter of its semantics and without pragmatic mediation, and that its hearer would need to engage in pragmatic-like reasoning about the speaker's referential intentions to understand its referent, the redundancy argument seems to overgeneralize.

My second aim will be to refine the original argument. I will argue that the redundancy observation does not convincingly apply only to these expressions whose semantic interpretation involves considering speakers' intentions as a result of the compositional process — that is, as a consequence of combining smaller semantic conventions.

References

Bach, K. (2017) Reference, Intention, and Context, in M. de Ponte and K. Korta (eds.) *Reference and Representation in Thought and Language*, 57–72. Oxford: Oxford University Press.

Heck, R. (2014). Semantics and Context-Dependence: Towards a Strawsonian Account. In Brett Sherman & Alexis Burgess (eds.), *Metasemantics: New Essays on the Foundations of Meaning*. Oxford University Press. pp. 327–364.

King, J. C. (2014) Speaker Intentions in Context, *Noûs*, 48/2: 219–37.

Radulescu, A. (2019). A Defence of Intentionalism about Demonstratives. *Australasian Journal of Philosophy* 97 (4): 775–791.

Smit, J. P. (2012). Why Bare Demonstratives Need Not Semantically Refer. *Canadian Journal of Philosophy* 42 (1):43–66.

Stokke, A. (2010). Intention-sensitive semantics. *Synthese* 175 (3):383–404.

Philosophy of Language
Seweryna Łuszczewska-Romahnowa on Clarity
Zuzana Rybaříková

University of Ostrava, Czech Republic | zuzana.rybarikova@osu.cz

Philosophy has always been considered to be an obscure and opaque discipline, full of ambiguous and complicated constructions. However, many contemporary philosophers stress the importance of a clear style in both philosophy and science and have a negative view of those who fail to adhere to this criterion (cf. Będkowski 2019b; Buekens & Boudry 2015). This claim is not new in philosophy. Already Heraclitus was called "obscure" by his contemporaries because of his unclear expressions (cf. Joll 2009, 85). In my talk, I will focus on the demand for a clear philosophical style formulated by the philosophers of the so-called Lvov-Warsaw School. In particular, I will discuss the reception of this demand by Seweryna Łuszczewska-Romahnowa.

Kazimierz Twardowski, the founder of the Lvov-Warsaw School, stressed that philosophy should be presented clearly. The clear and comprehensible presentation of philosophy was also one of the reasons for his success as an educator (see Będkowski 2019a, 90-91). He also presented his views on clear and unclear philosophical style in his eponymous essay "O jasnym i niejasnym stylu filozoficznym" [On Clear and Unclear Philosophical Style] (see Twardowski 1919), which became the Lvov-Warsaw School's manifesto. However, there was neither a generally accepted definition of clarity nor a consensus on how unclear philosophical texts should be treated (see Brożek et al. 2020, 194). While the logical branch of the school, especially Jan Łukasiewicz and Stanisław Leśniewski took a radical stance on unclear philosophical style, i.e. they rejected all texts that they considered obscure, other members of the school pointed out the disadvantages of this position (see Czeżowski 1969). Łuszczewska-Romahnowa was one of them.

Jacek Jadacki (2018, 26-27) presented several perspectives from which to consider textual clarity and unclarity. First, a text may be unclear because it deals with a complicated philosophical (or scientific) issue that cannot be clearly presented. Second, a text may be unclear because the author himself is not thinking clearly. Thirdly, an unclear text may result from the author thinking clearly but failing to articulate what he is thinking. However, Jadacki does not mention a case that is discussed in detail by Buekens and Boudry (2015). The possibility is that the problem is not too complex, the author thinks clearly and would be able to present his or her ideas in this way, but for various malicious reasons decides to write his or her text in an unclear way. This latter approach is called obscurantism. The members of the Lvov-Warsaw School did not deal with it. However, I would like to use the

example of Łuszczewska-Romahnowa's approach to show that their approach to the question of clarity in philosophy can also contribute to the current discussion of obscurantism.

Łuszczewska-Romahnowa introduced the question of clarity several times in her work. The focus of my presentation is on two of them, namely the question of clarity in Twardowski's work and the question of ambiguity in the language of science (Łuszczewska-Romahnowa 1967; 1977; 1948). On the one hand, Łuszczewska-Romahnowa (1967, 160-161) appreciated the clarity of style in philosophy and pedagogy and considered the demand for clarity as one of the great contributions of the Lvov-Warsaw School to Polish philosophy. On the other hand, she stressed that the quick discarding of obscure texts could do more harm than good. She argued that in philosophy (as in science) even vaguely formulated theories can yield valuable results (see Łuszczewska-Romahnowa 1948/1979, 153-155; 1967, 161-162).

As I have already mentioned, the members of the Lvov-Warsaw School did not concern obscurantism in the discussion over clarity. They did not reflect on the possibility that the author wrote obscurely on purpose; rather, following Twardowski, they argued that an obscure philosophical style is often a sign of obscure thinking. And since, according to them, obscurely thinking authors do not deserve to have the reader bother with their texts, they refused to read them. Yet their practice of avoiding unclearly written texts may have acted as an antidote to the dangers of obscurantism, both in the texts they wrote and in their acceptance and appreciation of the texts of other philosophers. However, Łuszczewska-Romahnowa's critique points out that it is not always wise to rush to dismiss all that is unclear and ambiguous. It also reminds us that an unclear style is not always a defect and that accusing an opponent of obscurantism should always be thoughtfully considered.

Bibliography

Będkowski, M. (2019a). „Jasnościowcy". O stylu naukowym Szkoły Lwowsko-Warszawskiej z perspektywy idei prostego języka (rekonesans). *Oblicza Komunikacji, 11*, 87–106.

Będkowski, M. (2019b). Nauczyć krytycznego myślenia i jasnej mowy. Postulaty krytycyzmu i jasności a sprawa tzw. logiki ogólnej. *Studia Semiotyczne*, *33*(2), 167–183.

Brożek, A., Będkowski, M., Chybińska, A., Ivanyk, S., Traczykowski, D. (2020). *Anti-irrationalism. Philosophical methods in the Lvov-Warsaw School*. Semper, Warsaw.

Buekens, F., & Boudry, M. (2015). The dark side of the loon. Explaining the temptations of obscurantism. *Theoria, 81*(2), 126–142.

Cześowski, T. (1969). Zarzut niejasności (przyczynek do teorii dyskusji). In: Cześowski, T. *Odczyty filozoficzne* [wyd. II]. Toruń: Towarzystwo Naukowe w Toruniu. 197–200.

Jadacki, J. (2018). Jeszcze o jasności myśli i mowy (w nawiązaniu do jednego z artykułów programowych Kazimierza Twardowskiego). *Filo-Sofija*, *40*(1), 23–39.

Joll, N. (2009). How Should Philosophy Be Clear? Loaded Clarity, Default Clarity, and Adorno. *Telos*, 146(Spring 2009), 73–95.

Łuszczewska-Romahnowa, S. (1948/1979). Wieloznaczność a język nauki. *Kwartalnik Filozoficzny* t. XVII/1948 s. 47–58. English: Ambiguity and the Language of Science, *Semiotics in Poland 1894-1969*. J. Pelc (red.), PWN, Warsaw, s. 184–195.

Łuszczewska-Romahnowa, S. (1967). Program filozofii naukowej Kazimierza Twardowskiego. *Studia Filozoficzne* 4, 154–168.

Łuszczewska-Romahnowa, S. (1977). Teoria wiedzy Kazimierza Twardowskiego. W: B. Skarga (red.), *Polska myśl filozoficzna i społeczna*. T. III. Książka i Wiedza, Warszaw, s. 86–125.

Twardowski, K. (1919). O jasnym i niejasnym stylu filozoficznym. *Ruch filozoficzny* V, 25–27.