DE GRUYTER

CONTEXT DEPENDENCE IN LANGUAGE, ACTION, AND COGNITION

Edited by Tadeusz Ciecierski and Pawel Grabarczyk

EPISTEMIC STUDIES
PHILOSOPHY OF SCIENCE, COGNITION AND MIND

Context Dependence in Language, Action, and Cognition

Epistemic Studies

Philosophy of Science, Cognition and Mind

Edited by
Michael Esfeld, Stephan Hartmann
and Albert Newen

Context Dependence in Language, Action, and Cognition

Edited by
Tadeusz Ciecierski and Paweł Grabarczyk

DE GRUYTER

ISBN 978-3-11-070207-1
e-ISBN (PDF) 978-3-11-070228-6
e-ISBN (EPUB) 978-3-11-070233-0
ISSN 2512-5168

Library of Congress Control Number: 2020944501

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Table of Contents

Paweł Grabarczyk and Tadeusz Ciecierski

Introduction

The phenomenon of context-dependence is so multifaceted that it is tempting to classify it as heterogeneous. It is especially evident in the case of the difference between context-dependence as understood in the philosophy of language and context-dependence as understood in the philosophy of mind. One of the aims of the present volume is to show that as varied as the phenomenon of context-dependence is, the similarities between its different manifestations are profound and undeniable. More importantly, as evidenced in a number of papers presented on the subsequent pages of this volume, a broad perspective on the phenomenon of context-dependence helps us to re-apply theories devised for one of the subfields of philosophy to the other subfields. Since the connections and analogies between many uses of contextualism may not be initially obvious, keeping an open perspective and the willingness to learn from the work of others may sometimes be crucial for finding new, satisfactory solutions.

The idea of the present collection of papers is the result of the first Context, Cognition and Communication conference which took place in Warsaw in the summer of 2016. The conference attempted to combine three perspectives—that of philosophy of language, that of philosophy of mind and that of philosophically oriented linguistics. We believe that the main idea behind the conference has been well reflected in the current collection.

The first five papers of the collection analyze context-dependence from the perspective of philosophy of language. In the first paper, Esben Petersen shows how classic philosophical puzzles can sometimes be solved by context-dependence. The paper discusses Moorean paradoxes related to the usage of words such as 'certain' or 'know'. As pointed out by Moore, utterances such as '*p*, but I don't know that *p*' seem infelicitous but explaining the reason why they sound 'off' to us is not easy. Petersen suggests that the solution to these paradoxes involves context-dependence, because the epistemic appropriateness of assertion is always evaluated against a tacit scale measuring epistemic certainty. This scale may provide a missing contextual parameter of our utterances, which determines when the epistemic support for a proposition is sufficient for assertion.

In "Meaning Holism and Contextualism[s]: Friends or Foes?" Filip Kawczyński discusses the connection between holism and contextualism—or rather the similarity between some of their problematic consequences. The well-known critical

Paweł Grabarczyk, IT University of Copenhagen, University of Lodz (pagrab@gmail.com)
Tadeusz Ciecierski, University of Warsaw (taci@uw.uni.pl)

https://doi.org/10.1515/9783110702286-001

argument against holism developed by Fodor and Lepore shows that one of the biggest problems of holistic theories is that they result in instability of meaning. If the meaning of every expression in language depends on the meaning of all of the other expressions, then any change of the meaning of any of the expressions necessarily changes the meanings of all expressions. A very similar objection can be made to contemporary contextual theories. The connection between holism and contextualism is especially visible in the case of 'local holism'—theories that apply the holistic principles to smaller units of language, such as theories or, especially, sentences. In these cases, holistic and context theories seem to claim a very similar thing—namely that the meaning of expressions depends on the larger unit they are a part of. This results in several well-know counterintuitive consequences for context theories. For example, if meaning is too sensitive to the context, then it is hard to imagine how genuine disagreement is even possible. Despite these similarities, Kawczyński argues that the connection between holism and contextualism is much weaker than it appears to be on the surface.

The aforementioned problem of genuine and apparent disagreements is the topic of Marian Zouhar's paper entitled "On the Nature of Non-Doxastic Disagreement about Taste". Zouhar starts by describing the difference between situations in which users manifest an evaluative attitude toward something and situations in which they express it. As pointed out by the author, some expressions, such as 'this chocolate is delicious' may contain double content—they can be manifestations and expressions at the same time. On the face of it, predicates of taste should not lead to disagreements, because they are always relativized to the user. This is not obvious and can be best illustrated by sentences about someone else's taste evaluations. Imagine a situation in which language users observe someone eating chocolate. One of the observers comments on it by saying that "this chocolate is delicious." The other one disagrees. It seems that in this case the disagreement can be genuine. Zouhar believes that this shows that taste predicates are context-sensitive—they function differently, depending on whether they are used in the first person, the third person, or if they are parts of more complex sentence structures.

A discussion of a somewhat similar problem can be found in "Why the Basic Problem Is Not a Problem" by Joanna Klimczyk. As pointed out by the author, the titular "basic problem" can be succinctly characterized in the following way: "it is impossible for one agent to stand in the normative relation to a proposition expressing another agent's activity." What it means in practice is that some normative claims can be interpreted correctly only when expressed in the first person, which makes them context-sensitive. This idea is incompatible with an influential reading of normative expressions devised by Mark Schroeder, who suggested that any normative content can be expressed by a sentence with an

ought operator. Klimczyk's solution to this problem is to show how some of the seemingly first-person only sentences can be re-interpreted from the third-person point of view.

In "Gettier Cases, Warranted Assertability Maneuvers, and the Fourth Condition" Tomasz Puczyłowski considers a powerful objection against contextualism presented in the form of a warranty assertability maneuver formulated by Keith DeRose. Puczyłowski assumes a very general strategy. If we are to believe that epistemic contextualism is, in fact, threatened by WAM, we have to trust WAM to be a good tool for contesting theories. If we show that, at least in some cases, the strategy does not work as intended, then its relation to epistemic contextualism should be reconsidered. Puczyłowski uses the test case of the Gettier problem and shows that WAM, in the form presented by DeRose, cannot be treated as decisive.

The next three papers consider context-dependence from the perspective of philosophy of mind. In "Self vs Other? Social Cognition, Extended Minds, and Self-Rule" Andrew Sneddon points out that the social side of human cognition seems to threaten our ability of self-ruling. If our behavior is always co-determined and dependent on the social context, then, even if we are still 'selves' we may not be autonomous. Sneddon shows that this fear can only be multiplied once we combine the idea of socially distributed cognition with the extended mind hypothesis. The idea of self-rule and autonomy requires us to discern between ourselves and our surroundings, but this demarcation seems to be impossible if the social context is to be treated as constitutive for cognition.

The connection between contextualism and human cognition is also the topic of the next paper—"Articulating Context-Dependence: *Ad Hoc* Cognition in the Prototype Theory of Concepts" by José V. Hernández-Conde. The author explores the connection between prototype theories of concepts and the radical contextualists' view according to which all concepts are contextually dependent. What it means in practice is that concepts are always instantiated or generated by the brain. The reason for this fluid nature of concepts is that brains, as well as the circumstances in which brains operate, change constantly. Hernández-Conde seeks to bridge the gap between prototype theories of concept and the contextual *ad hoc* approach by suggesting a dual nature of concepts. On the one hand, concepts can be understood as storage; on the other hand, they can be understood as real-time instantiations. Analogously to the relation between literal meaning and its contextual parameters, stored concepts do not determine the acts of categorization but are supplemented with additional variables available at instances.

The next paper, namely "Success and Knowledge in Action" by Markus Kneer, starts by criticizing Anscombe's thesis that knowledge is necessary for action to be classified as intentional. Kneer proceeds with a presentation of his

own account of the relation between knowledge and action. Following MacFarlane, he shows that the addition of the context of the assessment of knowledge ascription enables one to escape the criticism of Anscombe's intuitions.

In the last paper of this section, "*De Re* Explanation of Action in Context: The Problem of 'Near Contraries' and Belief Fragmentation", Sean Crawford shows how the application of contextualism helps us to retain *de re* descriptions of agents' psychological states without attributing contradictory beliefs to these agents.

The third group of papers consists of practical applications of philosophical context theories to specific domains. The first of these papers, Maria Caamaño-Alegre's "The Role of Presuppositions and Default Meaning in Framing Effects" uses contextualism to explain framing effects. A particular case of framing effects Caamaño-Alegre discusses relates to value statements which are sometimes used in polls. It can be shown that different but extensionally equivalent paraphrases of the same statement can sometimes result in very different reactions from the respondents. This phenomenon can be either explained as a manifestation of our cognitive limitations or as a cognitive strategy used during the interpretation. Caamaño-Alegre suggests that what differs here is the typical or most likely context these formulations are associated with.

A very different application of contextualism can be found in "Transcending the Situation: On the Context-dependence of Practice-based Cognition" by Rasmus Gahrn-Andersen. Just as Andrew Sneddon did in his paper, the author looks at distributed cognition as a form of context-dependence. As Gahrn-Andersen points out, distributed cognition theories typically focus on synchronic relations between agents—they focus on the influence of previous situations on the current actions. The author suggests expanding this view with a diachronic perspective. He shows a concrete application of this idea through the case study of a Danish company working on leakage detection. He shows that actions performed by its employees are best understood as heavily contextualized, as they build their identity on their relation to the actions of other employees, previous actions of the same and other agents (both human and artificial) as well as intended future actions.

The last paper in the collection shows a possible application of contextualism for empirical linguistics. In "Indirect Relations and Frames: Coreference in Context" Magdalena Zawisławska and Maciej Ogrodniczuk analyze the phenomenon of indirect reference (for example reference by association) using a case study of the Polish language. Their work is based on the existing study of the corpus of Polish language and, in particular, on some of the problems revealed in the annotation phase of the development of the corpus. The authors propose an

original typology designed for cases not covered by competing approaches (which are typically grouped together as ‘others’).

Esben Nedenskov Petersen

Moorean Paradoxes, Assertion, and Certainty

Abstract: While Moorean paradoxes with 'know' and epistemic 'certain' are distinct, sentences of both types are infelicitous to assert. Jason Stanley and Timothy Williamson both purport to explain these data based on their respective accounts of epistemically appropriate assertion. Stanley claims that his Epistemic Certainty Norm of Assertion provides a unified account of the observed infelicity, while Williamson explains it by supplementing the Knowledge Norm of Assertion with further assumptions about the relation between knowledge and certainty. In this paper, I argue that neither of these explanations succeeds. I then suggest that a unified account of Moorean paradoxes with 'know' and epistemic 'certain' may be provided by modifying Stanley's Epistemic Certainty Norm.

1 Introduction: Moorean Paradoxes and Norms of Assertion

Moorean paradoxes play a central role in the discussion about the epistemic requirements on appropriate assertion. It is commonly assumed that these norms are the key to explaining why Moorean paradoxes are infelicitous (DeRose 2003; Kvanvig 2009; Turri 2011, 2013; Unger 1975; Weiner 2005; Williamson 2000, 2009). Among these constructions, Moorean paradoxes with 'know' usually receive the most attention:

(1) *P* but I don't know that *p*.

For example, 'Paris is the capital of France, but I don't know that Paris is the capital of France.'

Assertions of this type sound wrong, but are clearly not contradictory. So their infelicity seems to have a pragmatic source. And since the infelicity is not limited to specific contexts, it is highly plausible that it reflects a general norm governing the practice of assertion.

Meanwhile, assertions with 'know', it seems, are not the only Moorean paradoxes relevant to accounts of assertion. The kinds of considerations which sug-

Esben Nedenskov Petersen, University of Southern Denmark (esben@sdu.dk)

https://doi.org/10.1515/9783110702286-002

gest that the infelicity of (1) owes to the presence of a general norm of assertion seem to apply with equal strength to analogous constructions with propositional 'certain':

(2) '*P* but it is not certain that *p*.'

For example, 'Scipio Africanus died in Liternum, but it's not certain that Scipio Africanus died in Liternum.'

Both Moore-sentences with 'know' and with 'certain' are infelicitous. So an account of assertion should not only explain why Moorean paradoxes with knowledge denials are infelicitous. It should also consider the datum that Moorean paradoxes with propositional 'certain' are infelicitous.

This paper discusses responses to this challenge. First, I consider Jason Stanley's (2008) proposal that his Epistemic Certainty Norm of Assertion explains the infelicity of both (1) and (2). Against this proposal I argue that Stanley's account of the infelicity of (1) does not succeed. The paper then proceeds to consider the treatment of (1) and (2) proposed by Timothy Williamson as part of his influential defense of a Knowledge Norm of Assertion, KN (Williamson 2000; 2009). But while KN easily accounts for the infelicity of (1), Williamson's proposal has problems explaining why assertions of (2) are infelicitous. After discussing the accounts proposed by Stanley and Williamson, I then tentatively suggest a way in which a unified account of (1) and (2) might be achieved by modifying Stanley's Epistemic Certainty Norm.

2 Stanley's Account of Moorean Paradoxes

Contrary to orthodoxy, Jason Stanley (2008) argues that a Certainty Account of Assertion, CA, should replace the Knowledge Account of Assertion, KA. According to this proposal, assertions must comply with an Epistemic Certainty Norm of Assertion, ECN, to be appropriate:[1]

ECN
Assert that *p* only if it is epistemically certain for you that *p*.

1 Stanley also posits a Subjective Certainty Norm of Assertion, SCN, which enjoins speakers only to assert propositions they are subjectively certain of. The conjunction of ECN and SCN form the Certainty Norm of Assertion, CN, which Stanley thinks governs the practice of assertion. Meanwhile, since SCN is inessential to Stanley's treatment of (1) and (2) I shall not discuss this norm further here.

Among the supposed advantages of this proposal, Stanley emphasizes ECN's ability to account for both of the aforementioned types of Moorean paradoxes. But while ECN elegantly explains the infelicity of (1), I will argue that Stanley's account of Moorean paradoxes with 'know' is flawed. Before looking at how Stanley explains why (1) and (2) sound strange, however, it is important to clearly distinguish varieties of propositional certainty from subjective certainty.

Subjective certainty refers to the type of certainty at issue when a subject is described as certain of something, as in 'I'm certain that dogs bark' (Audi 2003; Klein 1998; Reed 2011; Stanley 2008). What makes a person S subjectively certain that *p* relative to a context *c* is something like being sufficiently confident in her belief that *p*, according to the standards in *c*, or having a sufficiently unshakeable belief that *p*. This means that the predicate of subjective certainty is non-factive. One may be completely confident or unshakeable in one's belief that *p* although *p* is false.

Propositional certainty is the type of certainty at issue with assertions that describe a proposition as being certain, e.g., 'It is certain that dogs bark.' Such certainty comes in two varieties, epistemic and metaphysical. A proposition *p* is a propositional certainty in the metaphysical sense iff *p* is a settled fact, as opposed to a contingent fact about the open future, for instance. In contrast, whether *p* is an epistemic propositional certainty for S in *c* depends on S's evidence for *p*. For *p* to be an epistemic propositional certainty for S her evidence for *p* needs to be sufficiently strong relative to a contextually set standard in *c* to somehow guarantee that *p*.[2] In contrast to expressions of subjective certainty, expressions of propositional certainty are factive on both their epistemic and metaphysical interpretation since *p* must be true both for the evidence to guarantee that *p* is true, and for *p* to be a settled fact.

Assertions of type (2) are infelicitous on both a metaphysical reading of 'certain' and an epistemic reading. Alternating between these two interpretations does not prevent the assertions from sounding odd. For present purposes, however, the constructions of primary interest are those where 'certain' is used epistemically, since both Stanley and Williamson are well-positioned to explain the infelicity of constructions where 'certain' gets a metaphysical interpretation.[3]

2 Note though that the distinction between evidential and metaphysical readings of 'certain' is not an entirely sharp one (DeRose 1998). For example, if we are sufficiently liberal about evidence, the trajectory of a flying baseball may serve as evidence that it will land in a particular place (DeRose 1998:74–79) but also settles where it will land.

3 Thus, assuming that only a settled fact may be epistemically certain or known, explaining why (2) is infelicitous on its metaphysical reading is straightforward for both Stanley and Williamson since this implies that asserting something epistemically certain or known entails

ECN easily explains why (2) sounds odd when 'certain' gets an epistemic reading. To be assertible according to ECN, the second conjunct of (2) must be true. But if the second conjunct is true, then *p* is not epistemically certain for the asserter. So, according to ECN, unless the first conjunct of (2) is inappropriate to assert, the second conjunct must be inappropriate to assert.

Stanley also claims that ECN explains why (1) sounds odd, despite acknowledging that epistemic certainty entails neither belief nor knowledge, i.e., that *p* may be epistemically certain for S although S does not believe or know that *p* (2008: 49). The crucial thought underlying his explanation of the oddness is the assumption that *p*'s being epistemically certain for S endows S with a disposition to acquire knowledge that *p* when the right circumstances obtain: "If a proposition is an epistemic certainty for a person at a time, then it does follow that the person is in a *position to know* that proposition. Being in a position to know a proposition is to be disposed to acquire the knowledge that that proposition is true, when one entertains it on the right evidential basis" (Stanley 2008: 49).

Stanley therefore suggests that when *p* is epistemically certain for S, simply entertaining the proposition will be enough for S to come to know *p*. Because of this relation, *p*'s being epistemically certain for S is supposed to imply that S's entertaining *p* in the way required to assert that *p* will ensure that S comes to know *p* (Stanley 2008: 49). By appealing to this assumption Stanley can offer an account of the oddity of (1) along the following lines. For its first conjunct to be appropriately assertible for S, *p* has to be epistemically certain for S. Therefore, if S entertains the proposition that *p* in order to assert it, her being in a position to know that *p* will ensure that she comes to know that *p*. This in turn will imply that the second conjunct of (1) must be false, which makes it inappropriate to assert, according to ECN, owing to the factivity of epistemic certainty.

However, as I will argue, this explanation does not withstand more careful scrutiny. It suffers from at least two problems.

3 Two Problems with Stanley's Explanation

The first problem with Stanley's account concerns the crucial assumption that if *p* is epistemically certain for S, then S is disposed to achieve knowledge that *p* by entertaining *p* based on the right evidence. The problem arises from familiar facts about the semantics of epistemic modals.

---

that the asserted proposition is metaphysically certain. Any assertion of a proposition which is not metaphysically certain will thus violate both KN and ECN.

When 'it is certain that *p*' uses 'certain' as an epistemic modal its behavior resembles that of other modal terms used epistemically, e. g., the modal auxiliary verbs 'must' and 'might', or the adjective 'possible'. On the canonical theory of epistemic modal expressions, developed in seminal work by Angelika Kratzer (1977, 1981, 1991, 2012),[4] these terms function as quantifiers over sets of possibilities in a modal base comprising the possibilities 'compatible with the pooled information' (von Fintel & Gillies 2008: 9) of a contextually specified group of agents.[5] In addition, context may also determine a ranking of the possibilities in the modal base depending on their proximity to a designated possible world referred to as the ordering source. The set of worlds that a modal quantifies over in a given context comprises the best ranked possibilities in the modal base, i. e., the possible worlds sufficiently close to the contextually specified ordering source. Epistemic necessity operators, including epistemic 'certain', function as universal quantifiers over such sets of possibilities. Roughly speaking, then, 'it is certain that *p*' is true in a context *c* iff *p* is true in all the best possible worlds compatible with the combined evidence of the group of agents relevant in *c*.

This in turn poses a problem for Stanley, since it suggests that 'it is certain that *p*' might be true relative to S's context *c* owing to a part of her personal evidence *e* which excludes the possibility that not-*p*, although she has not realized that *e* excludes this possibility. The example below confirms this prediction.

INVESTIGATOR: Anna is one of the investigators working on a murder case. After months of investigation Anna and the rest of the group agreed that Smith was the culprit and should be arrested on suspicion of murder. Two weeks after the arrest of Smith, however, Anna is still thinking about the case. One night, going over the evidence in her thoughts for the umpteenth time, she suddenly realizes that some of the evidence that the police have had for months actually rules out the possibility that Smith is the murderer. She goes through her reasoning again to make sure, then calls the inspector on the phone, and says: 'We've got the wrong guy chief. It turns out that it's actually been certain the whole time that Smith is not the murderer. We just didn't see it…'

4 For references to a broadly Kratzerian view of modality as the standard, canonical or leading linguistic theory of modality, see, among others, (von Fintel & Gillies 2008, 2011; Papafragou 2006; Portner 2009; Rett 2012; Yalcin 2011). Obviously, a defense of Kratzer's approach against competing views cannot be undertaken here. But see (Portner 2009) for a discussion and a helpful overview.

5 The agents may even include non-biological information stores, e. g., such things as ship logs (Hacking 1967) and unopened letters (DeRose 1991).

The central data point here is that when Anna says 'it's actually been certain the whole time that Smith is not the murderer' her utterance seems fine. Yet, on its most natural reading her assertion implies that there was a time *t* such that 'it was certain at *t* that Smith is not the murderer' is true relative to the context of Anna's phone call to the inspector, even though at *t* Anna and everyone else thought that Smith was the murderer. Her realization that Smith is innocent does not owe to recently acquired evidence but to her recent recognition of the implications of her old evidence. So Smith's innocence seems to have been certain for Anna at a time where she did not believe him innocent, and owing to evidence that she had already considered repeatedly at that time.

To conclude that the example raises a problem for Stanley's explanation, however, some further issues need to be considered. First, as mentioned above, 'it is certain that *p*' has both epistemic and metaphysical readings (Condoravdi 2002; DeRose 1998). But it is only on epistemic readings that a proposition's certainty depends on the evidence for it.[6] On a purely metaphysical reading it is enough for a proposition to be certain that it represents a fact which is somehow settled, e. g., settled in the same sense as facts about the past. So one might be tempted to question whether Anna's assertion actually concerns epistemic certainty, as opposed to metaphysical certainty. But this worry is insubstantial. Since facts about the past obviously settle who the murderer is, informing the chief investigator that these facts are metaphysically certain would be a very strange thing for Anna to do. Pragmatic restrictions on conversational relevance thus dictate the epistemic reading of her assertion.

Secondly, the gradability of 'certain' gives the term a particular type of context-sensitivity that it shares with gradable expressions in general, which should be considered here. A gradable predicate 'F' has a semantic link to a scale measuring the degree to which the property ascribed by 'F' is instantiated, while a contextually fixed threshold determines the degree to which something has to instantiate the relevant property to satisfy the predicate 'F' relative to a particular context of utterance (Kennedy 2007). The term 'flat', for example, has a semantic link to a scale measuring degree of flatness, and how flat something has to be to satisfy 'flat' depends on the context. So, presumably, how certain *p* must be, or how strong the evidence for *p* has to be for *p* to be epistemically certain depends on the context. One might then suggest that while the standards of certainty in Anna's context of utterance *c* may be low enough for 'it was certain that Smith is innocent' to be true in *c*, there may not have been any previous time where she

6 At least insofar as these can be clearly distinguished. As explained in footnote 2, however, this might not always be the case.

was in a context *c** such that 'it is certain that Smith is not the murderer' was true relative to *c**. Such a proposal would be poorly motivated, however, since there is nothing to indicate that the standards for certainty are any lower than usual in the context of Anna's phone call. There is no reference to any mechanism which could plausibly be held responsible for a shift toward particularly lenient standards of certainty in the example's context. So nothing precludes stipulating that she was in a context with identical standards for certainty at some prior time where she falsely believed Smith to be the murderer.

Thirdly, the epistemic certainty of *p* relative to some context *c* plausibly requires that it is possible to come to know that *p* based on the contextually relevant group's evidence by carrying out a practicable investigation (Hacking 1967).[7] But this condition on epistemic certainty does not rescue Stanley's assumption. There clearly was a practicable way open to Anna in which she could come to know that *p* based on her evidence. Indeed, since she has found out that Smith is innocent by using her cognitive capacities it was *ipso facto* within her epistemic reach to come to this conclusion based on her evidence. That she did not connect the dots earlier obviously does not imply that she could not have.

INVESTIGATOR thus reveals that there is an exception to the generality of the purported conceptual relation between epistemic certainty and the disposition to acquire knowledge. Letting '*p*' stand for the proposition that Smith is not the murderer, and using 'S' to abbreviate 'Anna', INVESTIGATOR shows that there could have been a situation in which S's evidence made *p* epistemically certain for S relative to her context of utterance *c*, while S did not believe that *p*, although she had contemplated whether *p* based on the evidence that made *p* epistemically certain for her relative to *c*.[8]

Further, by altering INVESTIGATOR slightly we could depict a situation where S's total evidence makes *p* epistemically certain for her, while she believes *p* based on a subset of her total evidence, such that it is epistemically certain for her that this evidence is too weak to make the belief knowledge. For example, instead of agreeing with the other detectives she might have had a hunch that Smith was innocent. Suppose then that her evidence guarantees that her hunches are not sufficiently reliable for beliefs based on them to qualify as knowledge, but that she still believed Smith innocent based on her hunch. In that case, her

7 See also DeRose 1991, Egan 2007, McFarlane 2011.

8 Here I am assuming that when Stanley says that epistemic certainty entails the disposition for S to acquire knowledge that p if she entertains p on the basis of the right evidence, he must have in mind the evidence which makes p epistemically certain for S in her context. Otherwise, the intended interpretation of this crucial passage becomes entirely unclear.

total evidence might make Smith's innocence epistemically certain, while the basis for her belief that Smith is innocent is such that it is epistemically certain that she does not know that Smith is innocent. It then follows that her assertion of 'Smith is innocent, but I don't know that Smith is innocent' would conform to ECN, although ECN is supposed to explain why no such instance of (1) is felicitous to assert. So, Stanley's explanation does not work.

Even if Stanley had been right about the entailment from epistemic certainty to being in a position to know, however, his account of (1) would have been doubtful. To see why, consider that Stanley describes being in a position to know as a disposition to acquire knowledge that *p* when one entertains the proposition based on the right evidence. The explanation thus depends on the assumption that a speaker asserting *p* considers *p* based on the right evidence, whichever way the right evidence is defined. But asserting that *p* does not require one to entertain the proposition that *p* based on any particular part of one's total evidence. So the manifestation conditions of the disposition Stanley describes need not obtain when *p* is epistemically certain for a speaker who entertains *p* in order to assert that *p*. It then follows that Stanley's questionable assumption does not rule out a situation where *p* is epistemically certain for S and S asserts that *p* without knowing that *p*. Thus, as explained above, while S's total evidence *e* makes it certain that *p* relative to her context, her occurrent belief that *p* may be based on a proper subset of her evidence which offers much less support for *p* than her total evidence.

With respect to Moore-sentences with 'know' I hence submit that their infelicity is not plausibly explained by Stanley's proposal. The question then is how a more convincing account of Moorean paradoxes might be developed. In the next section, I consider two answers to this question which are based on the Knowledge Norm of Assertion, and argue that both are flawed. I then go on to suggest a way to modify ECN to explain why both (1) and (2) sound wrong.

4 Two Explanations Using the Knowledge Norm

There might seem to be an easy way to account for the infelicity of (1) and (2) by combining ECN with KN. To remedy Stanley's problem, one might propose simply to add the familiar knowledge norm of assertion to ECN as a further requirement on appropriate assertion. Since the knowledge norm, KN, enjoins speakers only to assert what they know, this proposal would let Stanley explain the infelicity of (1) by using the same template as he uses to explain why (2) sounds odd. But while this straightforward approach would fit the data, appealing to the conjunction of ECN and KN to account for (1) and (2) would be to forego the theo-

retical unity of a single principle emphasized by Stanley as a main virtue of his original proposal (Stanley 2008: 49).

Insofar as explanatory unity is a priority, Timothy Williamson's (2000; 2009) explanation of (1) and (2) thus seems to be a more attractive alternative. Although Williamson agrees that knowledge does not entail certainty, he still thinks that the infelicity of Moorean paradoxes with 'know' and 'certain' is explained by KN. The reason why these constructions sound paradoxical, he argues, is not that appropriate assertion generally requires certainty as well as knowledge, but rather that speakers are reluctant to let the context-sensitive standards for certainty come apart from the standards for knowledge (Williamson 2000: 254; 2009: 344–5). As a result, their typical tendency will be to treat *p*'s not being epistemically certain for S as implying that S does not know *p*, which entails that a speaker asserting *p* while denying *p*'s certainty will be interpreted as violating KN.[9]

The problem with this proposal is that it finds nothing wrong with assertions of (1) and (2) in circumstances where the standards of knowledge and epistemic certainty do come apart. If Williamson's account is right, there should be situations where assertions of (2) are perfectly fine. Yet, such situations are missing, and Williamson's own purported example of one is unconvincing. To show that KN prevails when the standards for knowledge and certainty come apart, he notes that '*p* and by Descartes' standards I cannot be absolutely certain that *p*' is fine to assert. But the interpretation of this example seems to misrepresent the function of 'by standards'-phrases.

Contrary to Williamson's implicit assumption, 'by standards'-phrases are not used to shift the contextual standards which determine the content of a context-sensitive predicate. 'By standards'-phrases rather have a function akin to a quotation device. Generally, they are used to indicate that some particular group applies a term in a particular way which is possibly atypical. Asserting 'Squares only have sides of equal length, but by the standards of Tim rectangles are squares', for example, does not shift the semantic content of 'square', but indi-

9 For the sake of accuracy, note that Williamson's proposal overlooks or disregards that there are conceptually distinct types of certainty. Strictly speaking, this makes it slightly unclear whether the hypothesized alignment with the standards for knowledge is supposed to hold for epistemic certainty only, psychological certainty only, or both. For the present purposes, however, I shall assume that he takes the suggested alignment to hold for ascriptions of epistemic certainty. Furthermore, Williamson's own remarks seem to favor this interpretation, since his discussion of standards of certainty includes reference to 'Absolute Cartesian standards' which are most naturally interpreted as absolute with respect to what they require of a subject's evidence.

cates that Tim uses this term in a particular, very unusual way.[10] So it is a questionable assumption that '*p* and by Descartes' standards I cannot be absolutely certain that *p*' corresponds to an utterance where 'certain' is used with the meaning assigned to it by the conventions of ordinary language. Williamson's example only shows that asserting *p* may be felicitous, although a speaker using '(absolutely) certain' in the same way as Descartes would not say that *p* is '(absolutely) certain'. He does not provide a convincing example of a situation where 'certain' retains its conventional meaning, while the separation of the standards for certainty and knowledge permits the assertion of a known proposition which is not epistemically certain. But without such cases Williamson's account of (2) remains questionable.

Of course, this alone does not suffice to establish that orthodoxy must be abandoned. But it shows that while KA has an obvious advantage with respect to (1), how it should deal with (2) is a much less straightforward issue. So if there is an alternative to the orthodox knowledge account which readily accounts for the infelicity of (1) and (2), this puts substantial pressure on the proponents of orthodoxy to explain why their approach should be preferred. It is interesting then that a modified version of ECN seems able to account for both (1) and (2) without sacrificing explanatory unity.

5 Another Certainty Norm

As mentioned above, I will propose that ECN may be amended to provide a unified explanation of why (1) and (2) are infelicitous. To improve Stanley's proposal, I begin by noting that the two objections to his treatment of (1) have a common source, namely that *p*'s being epistemically certain for S does not ensure that S holds the belief that *p* based on the evidence which makes the proposition certain for her. This observation suggests a natural way to improve ECN. The observed problem can be avoided by adding a requirement to the effect that what is asserted must be believed on the right basis.

To provide a certainty account of assertion which explains the infelicity of (1), I therefore propose to replace ECN with the Epistemic Propositional Certainty Norm, EPCN (Petersen 2019: 4694):

> EPCN
> Assert that *p* only if your belief that *p* is based on evidence which makes it epistemically certain for you that *p*.

10 For these observations regarding 'by standards'-phrases, see Petersen 2019.

This norm, in conjunction with the highly plausible assumption that believing *p* based on certainty level evidence suffices for knowing *p,* explains the oddness of both (1) and (2) straightforwardly. If believing *p* based on certainty level evidence entails knowing *p*, then a speaker asserting *p* must know that *p* in order to comply with EPCN. So EPCN plausibly implies that a speaker must only assert *p* if she knows *p* and *p* is epistemically certain relative to her context.

Evidence for the assumption that believing *p* based on certainty level evidence for *p* entails that knowing *p* comes primarily from the apparent absence of any counterexamples to the proposed entailment. If the entailment holds, it is easy to explain why there are no cases where someone has a belief based on certainty level evidence yet fails to know. There is also a more detailed rationale for the assumption, however, which specifies why each of the standard conditions on knowledge that *p* seems to be met by a speaker who satisfies ECN* with respect to *p*.

First, epistemic certainty is factive. So if S's evidence makes *p* epistemically certain for her, then *p* must be true. Secondly, since S's evidence for *p* needs to be quite strong in order for *p* to be epistemically certain for S, it seems safe to also assume that S will be justified in believing *p* if she believes *p* based on certainty level evidence. There are even conceptions of certainty which take epistemic certainty that *p* to require that one has the highest possible degree of evidence for *p* (Chisholm 1976; Firth 1967; Reed 2011; Russell 1948). Furthermore, saying of someone that she believes *p* based on evidence which makes *p* epistemically certain for her, while denying that her belief that *p* is justified, would sound very strange. The simplest available explanation is that the source of this infelicity is that such an assertion would be a contradiction. Thirdly, the level of evidence needed for *p* to be epistemically certain for S seems to prohibit the kind of epistemic luck involved in Gettier scenarios. Indeed, the epistemic certainty of a proposition is typically thought to require that the evidence somehow ensures, or guarantees, that the proposition is true (Firth 1967; Lewis 1929; Reed 2011). Therefore, if S's belief that *p* is based on evidence which makes *p* epistemically certain for her, then surely it is not merely a fortunate coincidence that her belief happens to be true. For example, suppose that S sees her colleague Jones driving a Land Rover and this leads her to believe that one of her colleagues owns a Land Rover, although Jones was actually just driving a Land Rover borrowed from her friend. And suppose also that S's belief is true by coincidence because her other colleague Smith secretly owns a Land Rover which he keeps hidden in his garage. Might we then still claim that the evidence S has makes it epistemically certain for her that one of her colleagues owns a Land Rover? The answer is, obviously, no. Gettierization seems to be incompatible with epistemic certainty.

It hence seems that if S believes that *p* based on evidence which makes *p* epistemically certain for her, she must satisfy the conditions on knowing that *p*. If the epistemology textbooks are right that S knows *p* if S has a justified, true belief that *p*, which is not gettierized, then S's complying with EPCN when she asserts *p* ensures that she knows *p* when she asserts.

Insofar as the infelicity of Moorean paradoxes with 'know' and epistemic 'certain' should be explained by an account of assertion, this favors an approach based on EPCN over both Stanley's certainty account and Williamson's proposal based on KN. While Stanley's explanation fails to show why Moorean paradoxes with 'know' are infelicitous without exception, Williamson's approach makes the seemingly false prediction that assertions of Moorean paradoxes with epistemic 'certain' may be felicitous. The account of (1) and (2) based on EPCN avoids both these problems.

Of course, this is not enough to conclude that a certainty account of assertion based on EPCN is preferable to both Stanley's account and KA. What this paper achieves is not to establish a particular norm of assertion but to add an important nuance to the understanding of how Moorean paradoxes bear on the debate about norms of assertion. Since the capacity to deal with the types of Moorean paradoxes discussed here is not the only parameter by which theories of assertion should be measured, advantages on other parameters might give overall preference to a rival of the account based on EPCN. Resolving this question, however, would require a comprehensive comparison of the present proposal and competing accounts. The paper's discussion suggests that such an undertaking may be worthwhile, but leaves it for future work to pursue.[11]

6 Conclusion

Properties shared by Moorean paradoxes with 'know' and epistemic propositional 'certain' suggest that an account of assertion should explain why both these types of construction are infelicitous. Jason Stanley purports to explain their infelicity based on an epistemic certainty norm of assertion. But the explanation falters with respect to Moorean paradoxes with 'know' owing to a questionable assumption about epistemic certainty and its relation to knowledge. A different account proposed by Timothy Williamson purports to explain the infelicity of both types of Moore-sentences based on the Knowledge Norm of Assertion. But this account predicts that Moore-sentences with epistemic 'certain' should

11 For some discussion related to this question, see Petersen 2019.

sometimes be appropriate to assert, while permissible assertions of this type are missing. In contrast, a modified version of Stanley's epistemic certainty norm appears to provide a basis for a unified account of Moorean paradoxes with 'know' and epistemic 'certain'. The paper's discussion thus suggests that this certainty norm deserves a place in future investigations of the epistemic norms governing the practice of assertion.

Acknowledgements

I am grateful to my colleague Nikolaj Nottelmann for helpful discussion of the paper's issues, and to Caroline Schaffalitzky and Søren Harnow Klausen for their valuable comments on a draft version of the paper. Thanks for helpful discussion are also due to the audience attending the presentation of the paper at the 1st Context, Cognition, and Communication Conference, University of Warsaw, 2016.

References

Audi, Robert. 2003. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. New York: Routledge.

Chisholm, Roderick. 1976. *Person and Object*. La Salle, IL: Open Court.

Condoravdi, Cleo. 2002. "Temporal Interpretation of Modals". In David Beaver et al (eds.) *The Construction of Meaning*. Stanford: CSLI Publications: 59 – 87.

DeRose, Keith. 1991. "Epistemic Possibilities". *Philosophical Review* 100: 581–605.

DeRose, Keith. 1998. "Simple 'Mights', Indicative Possibilities, and the Open Future". *Philosophical Quarterly* 48: 67–82.

DeRose, Keith. 2002. "Assertion, Knowledge and Context". *Philosophical Review* 111: 167 – 203.

Egan, Andy. 2007. "Epistemic Modals, Relativism, and Assertion". *Philosophical Studies* 133: 1 – 22.

von Fintel, Kai and Gillies Anthony. 2008. "CIA Leaks". *Philosophical Review* 117: 77–98.

von Fintel, Kai and Gillies Anthony. 2011. "'Might' Made Right". In Andy Egan and Brian Weatherson (eds.), *Epistemic Modality* New York: Oxford University Press.

Firth, Roderick. 1967. "The Anatomy of Certainty". *Philosophical Review* 76: 3–27.

Hacking, Ian. 1967. "Possibility". *Philosophical Review* 76: 143 – 168.

Kaplan, David. 1989. "Demonstratives". In Joseph Almog, John Perry, and Howard Wettstein (eds.), *Themes from Kaplan*. Oxford: Oxford University Press.

Kennedy, Chris. 2007. "Vagueness and Grammar: the Semantics of Relative and Absolute Gradable Adjectives". *Linguistics and Philosophy* 30: 1 – 45.

Klein, Peter. 1998. "Certainty". In Edward Craig (ed.) Routledge Encyclopedia of Philosophy vol. 2. Routledge: London: 264 – 267.

Kratzer, Angelika. 1977. "What 'Must' and 'Can' Must and Can Mean". *Linguistics and Philosophy* 1: 337 – 355.
Kratzer, Angelika. 1981. "The Notional Category of Modality". In Hans-Juergen Eikmeyer and Hannes Rieser (ed.) *Words, Worlds, and Contexts. New Approaches to Word Semantics*. Berlin: Walter de Gruyter & Co.
Kratzer, Angelika. 1991. "Modality". In Arnim von Stechow and Dieter Wunderlich (eds.), *Semantics: An International Handbook of Contemporary Research*. Berlin: Walter de Gruyter & Co.
Kratzer, Angelika. 2012. *Modals and Conditionals*. Oxford: Oxford University Press
Kvanvig, Jonathan. 2009. "Assertion, Knowledge, and Lotteries". In Patrick Greenough and Duncan Pritchard (eds.). *Williamson on Knowledge*. Oxford: Oxford University Press.
Lewis, Clarens Irving. 1929. *Mind and the World Order*. New York: Dover.
MacFarlane, John. 2011. "Epistemic Modals Are Assessment-Sensitive". In Andy Egan and Brian Weatherson (eds.). *Epistemic Modality*. New York: Oxford University Press.
Papafragou, Anna. 2006. "Epistemic Modality and Truth Conditions". *Lingua* 116: 1688 – 1702.
Portner, Paul. 2009. *Modality*. Oxford: Oxford University Press.
Petersen, Esben. 2019. "A Case for a Certainty Norm of Assertion". *Synthese* 196(11). 4691 – 4710.
Reed, Baron. 2011. "Certainty". *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition). Edward N. Zalta (ed.). http://plato.stanford.edu/archives/win2011/entries/certainty/>.
Rett, Jessica. 2012. "On Modal Subjectivity". *UCLA Working Papers in Linguistics*, *Papers in Semantics*. 16: 131 – 150.
Russell, Bertrand. 1948. *Human Knowledge: Its Scope and Limits*. New York: Simon and Schuster.
Stanley, Jason. 2008. "Knowledge and Certainty". *Philosophical Issues*. 18, 33 – 55.
Turri, John. 2011. "The Express Knowledge Account of Assertion". *Australasian Journal of Philosophy* 89: 37 – 45.
Turri, John. 2013. "The Test of Truth: An Empirical Investigation of the Norm of Assertion". *Cognition* 129: 279 – 291.
Unger, Peter. 1975. *Ignorance: A Case for Skepticism*. Oxford: Oxford University Press.
Weiner, Matt. 2005. "Must We Know What We Say?". *Philosophical Review* 114: 227 – 51.
Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.
Williamson, Timothy. 2009. "Replies to Critics". In Patrick Greenough and Duncan Prichard (eds.) *Williamson on Knowledge*. Oxford: Oxford University Press.
Yalcin, Seth. 2011. "Nonfactualism about Epistemic Modality". In Andy Egan and Brian Weatherson (eds.) *Epistemic Modality*. Oxford: Oxford University Press: 295 – 332.

Filip Kawczyński

Meaning Holism and Contextualism[s]: Friends or Foes?

Abstract. In the paper I give arguments to justify the thesis that—contrary to common opinion—there is no strong theoretical relation between Meaning Holism and Contextualism. I distinguish two holistic accounts: Global Holism and Local Holism, and I investigate what are the relations between them and various versions of Contextualism, namely Eliminativism, the Wrong Format View, and Moderate Contextualism. I show that: (1) Local and Global Holism are compatible with almost all (with merely one exception) versions of Contextualism; (2) Holism (in any version) neither entails nor is entailed by any version of Contextualism. Thus, I argue that when it comes to the issue of context-sensitivity, a meaning holist, either global or local, has a variety of options to choose between, and vice versa—being a contextualist of this or that kind one can choose to accept either Local or Global Holism, to accept both, or to avoid both. Except showing that the relation between Holism and Contextualism is weak, in the paper I also sketch several possible ways of modifying these theories to make them theoretically closer or more distant.

1 Introduction: Instability of Meaning Holism and Contextualism[1]

Since the publication of the famous book by Jerry Fodor and Ernest Lepore (1992), Meaning Holism has—to put it mildly—lost its appeal among philosophers of language. The way Fodor and Lepore argue against Meaning Holism (hereafter: 'MH') echoes the opinion expressed by Fodor a few years earlier, namely that MH "really is a *crazy* doctrine" (Fodor 1987:60). What is claimed by Fodor and Lepore to be especially crazy and nerve-racking about MH is its alleged *instability:* if the meaning of a single expression depends on the meanings of all other expressions in a given language, then any change in any mean-

1 The work on this paper has been funded by National Science Center, Poland, grant under award number 2018/31/D/HS1/03745. I would like to thank Joanna Odrowąż-Sypniewska and Tomasz Zyglewicz for their valuable comments on the draft of this paper.

Filip Kawczyński, University of Warsaw (f.kawczynski@uw.edu.pl)

https://doi.org/10.1515/9783110702286-003

ing implies some changes in all the other meanings. In such a case, language can hardly be said to be *systematic* or *stable*. And if language is such an unstable and wobbly matter, it is quite difficult to imagine how it is possible for it to work with any degree of efficiency.

This kind of objection is commonly raised not only against MH, but also against views belonging to the contemporary mainstream in the philosophy of language—i.e. against various versions of Contextualism. Roughly speaking, Contextualism is the view that the semantic value of an expression can be partially determined by s*trong pragmatic effects*. It means that fixing the standard time-place-speaker parameters is sometimes not enough to determine semantic value and thus some additional contextual factors must be taken into account. According to opponents of this theory, it makes meanings *too* sensitive to contextual factors and leads to the general unwanted consequence that any systematic theory of language is impossible.

More specifically, these (allegedly) *unstable* theories are accused of leading to several problems. The first of them is called the problem of the possibility of *genuine disagreement*. If MH / Contextualism is true, then it is highly unlikely that two people mean the same by 'p', since its meaning depends on too many other meanings in one's idiolect (in the case of MH) or is too sensitive to contextual factors and thereby changes too rapidly to plausibly facilitate communication (in case of Contextualism). Thus, when one asserts 'p', while some other person denies it (i.e. states '¬p'), no genuine disagreement arises between the speakers because they use 'p' with different meanings. The extension of that problem—again applicable to both MH and Contextualism—is the problem of the possibility of *changing one's mind*, which is very close to the genuine disagreement issue. In this case, however, what is concerned is not two different speakers but one person expressing some beliefs at two different moments. If 'p' at one moment has a different meaning than 'p' at some later time and the person utters '¬p' at that later time, it should not be considered denying her earlier belief. And that stands against our common intuitions regarding such situations. The other problem concerns the *possibility of communication* at all. If MH or Contextualism imply that it is almost impossible for two people to express *the same thing* by uttering 'p', communication seems to be impossible altogether, while obviously we observe that it is not only possible but that it works quite fine. Hence, it is often said that accounts like MH and Contextualism entail the view that communication is a kind of miracle.[2]It seems that the above instability-track has led to

2 For a discussion on the instability:—of MH, see e.g. Bilgrami (1998), Block (1994), Brandom (2000; 1994), Churchland (1993:668 – 672), Field (1977), Fodor (1987), Fodor and Lepore (1992:

putting MH and Contextualism in the same pigeonhole labelled "crazy relativistic doctrines."[3]

In this paper, I argue that, contrary to common opinion, the ties of kinship between MH and Contextualism are much weaker than they may appear at first glance—in doing so, I hope to set MH in a new light and thus take a step toward revitalising it. In the next section I concisely characterize both theories in question, and I also provide the distinction between Global Holism and Local Holism. The third section is devoted to Global Holism, while the fourth addresses Local Holism. In these two sections, I try to set out what the relations are between these theories and three variants of Contextualism: Moderate Contextualism, Eliminativism and the Wrong Format View. What I aim to do in particular is, on the one hand, to answer the question concerning the compatibility of the discussed theories and, on the other, to investigate if a given version of MH entails or is entailed by any version of Contextualism. The entailment should be understood here strictly as a logical relation between the main theses of the theories. In other words, I attempt to discover if being a holist (of this or that kind) obliges one to accept a version of Contextualism, and vice versa. The final section includes a conclusion and some closing remarks.

2 Characteristics of the Theories

2.1 Meaning Holism

Defining and defending "a crazy theory" is always inconvenient. What makes it even worse in our case is that holism, as Christopher Peacocke (1997: 227) points out, is "a doctrine which is attributed to all Quine, Putnam, Davidson, Rorty, Gadamer and Heidegger"—which, as one must admit, is a remarkably diverse ensemble of thinkers. In turn, MH belongs to those unlucky philosophical theories that have not earned a satisfactory definition as yet. In the principal source of philosophical definitions, which I suppose is The Stanford Encyclopedia of Philosophy, Henry Jackman characterizes MH in the following manner: "The label 'meaning holism' is generally applied to views that treat the meanings of all of the words in a language as interdependent. Meaning holism is typically con-

17–22; 2002), Harman (1993), Jackman (1999, 2017), Lormand (1996), Pagin (1997, 2006), Rovane (2013).—of Contextualism, see e. g. Cappelen and Lepore (2005: ch. 8), Kölbel (2004), Recanati (2010: 6–10), Stanley (2000, 2002, 2005).

3 Michael Devitt was so terrified of MH that he called it a theory that threatens "life as we know it" (Devitt 1996: 10).

trasted with *atomism* about meaning (where each word's meaning is independent of every other word's meaning), and *molecularism* about meaning (where a word's meaning is tied to the meanings of some comparatively small subset of other words in the language [...]" (Jackman 2017: § 1).[4]

As we can see, there is not much about MH itself here and more about its rival theories. Peacocke delivers a more detailed account that he believes is a general definition of MH that all its adherents would agree on: "The meaning of an expression depends constitutively on its relations to all other expressions in the language, where these relations may need to take account of such facts about the use of these other expressions as their relations to the non-linguistic world, to action and to perception" (Peacocke 1997: 227).

Both above formulations—which are representative of the paradigmatic way of defining MH—concern something that I would like to call [*Meaning*] *Holism as a Principle*, which can be shortly formulated as:

> **(H-Principle)** The meaning of a single expression depends on the meanings of all other expressions in a given linguistic system.

This principle ought to be sharply distinguished from holism regarded as a possible *theory of meaning*. By a theory of meaning I understand an account that is supposed to deal with problems such as how meanings even come about in the world, what their ontological status is, or more generally what the nature of meaning is. Holism as a Principle (hereafter: 'H-Principle') does not address any of these issues. Roughly speaking, H-Principle makes a point about how meanings behave and interact and not about what they essentially are. Obviously, an analysis of something's behavior can tell us much about the constitution of the thing in question; however, it seems that the holistic thesis about the interaction of meanings is largely independent of particular answers to the question of the nature of meaning itself. As far as I am concerned, H-Principle can be reconciled with accounts of meanings as illocutions, meanings as inferential roles, meanings as uses, meanings as references (in Davidson's sense), stimulus mean-

4 Peter Pagin (2006) very accurately identifies several problems that arise from attempting to define MH in terms of determination or interdependence of meanings. However, I believe that those problems should be in fact regarded as problems for holism itself, i.e. as a challenge for making sense of holism in general, not merely as methodological issues with defining it. Notwithstanding this, it seems that it is the very idea of meaning interdependence that underlies all holistic semantic theories; rejecting that, in my opinion, is tantamount to excluding holism from the realm of possibly correct theories of language. In this paper I do not aim to directly defend MH, therefore I pass over the issues in question.

ings (if they are meanings at all) and probably even ideal meanings (from the Fregean 'third realm').

Since H-Principle is not a theory but a principle, I think it can be left in the form given above, without further pinpointing and with a dose of generality that principles are expected to bear. However, I believe that one additional step has to be taken. Namely, I would like to distinguish between applying H-Principle to *whole languages* and applying it to particular *units of language*. I call the former approach Global Holism and the latter Local Holism.[5]

2.1.1 Global and Local Holism

Global Holism (hereafter: 'GH') is the result of combining H-Principle with the Quine-inspired[6] claim that *meaning is attributed to a whole language and not to linguistic items in isolation*. As such, GH concerns meanings as occurring outside or before any particular context, so it can be said that GH deals with *literal meanings*.

Local Holism, on the other hand, is the result of applying the holistic paradigm to chosen units of language, e. g. individual sentences, utterances, speech-acts, theories, discourses and so on. The meaning of every single expression involved in a given unit depends on the meanings of all other expressions involved.[7] What Local Holism is interested in is the behavior of meanings in par-

5 'Global Holism' and 'Local Holism' appear in the literature (e. g. see: Peacocke 1997; Penco 2001) from time to time, however, without fixed meaning or reference; some of those uses are quite close to mine, others are not.

6 What I have in mind here is obviously the Duhem-Quine thesis: "the unit of empirical significance is the whole of science" (W. V. O. Quine 1951: 41).

7 You can think of two major interpretations of so defined Local Holism. According to the first, i. e. the molecularist reading, meanings involved in separate units are not interdependent. For instance, if a physical theory and a biological theory are taken to be such units, local holists claim that the meanings of all physical terms depend on each other and the meanings of all biological terms depend on each other, yet the meaning of no physical term depends on any meaning of a biological term, and conversely. The other interpretation, which may be called strongly holistic, assumes that there can be interdependence of meanings between words belonging to units that can be in a way separated. Thereby several variants of Local Holism may be distinguished, with regard to how many of such interdependencies are allowed. In the strongest version—according to which all units are somehow connected—this amounts to a view that resembles GH; however, it must be remembered that Local Holism is about meanings-in-contexts (or speaker's meanings), as opposed to GH, which is about literal meanings. In what follows, I ignore that further classification of different versions of Local Holism because it does not affect the results of my investigation of what the relation between Local Holism and Contextualism is.

ticular contexts. Thus, Local Holism crosses paths with Contextualism; however, to make these two views genuinely comparable some further constraints should be imposed on Local Holism. Namely, one of the two main principles of Contextualism—see point (c1) below—considers propositions expressed by *sentences* in contexts. Hence, I am going to take into account a variant of Local Holism that captures relations between meanings within a sentence as used in a context:

> **(LH)** The meanings of all simple expressions appearing in a sentence uttered in a context are determined in a holistic way.[8]

A combination of GH and LH delivers an account that offers both the general perspective of looking at language as a whole as well as analyzing particular linguistic acts. However, although quite easy to reconcile, the theories in question are in fact logically independent.

On the one hand, GH does not entail LH: global holists are not committed to the view that within a particular sentence any meaning-formatting even happens. They can assume that literal meanings determined by the semantic structure of the whole language are what simple expressions contribute to the meanings of complex expressions. Then there is no point in saying that meanings within a sentence are determined holistically (or any other way). Furthermore, even when it is admitted that literal meanings are adjusted when appearing in a sentence, global holists can adopt a non-holistic approach for explaining that. In other words, global holists can claim that literal meanings are 'defined' holistically, while what happens with them in specific contexts is characterized correctly in e. g. an atomistic framework.[9]

There is no entailment the other way around either: it is not the case that LH entails GH. Analogously to the previous case, a local holist does not even have to be concerned about what is going on at the level of language as a whole, since LH is a view concerning the behavior of meaning within particular sentences. Nevertheless, LH can be readily reconciled with e. g. an anti-holistic atomistic view on literal meanings to the effect that literal meanings are determined in a way that does not assume mutual influences; however, when being modified

---

However, as a historical point it may be noticed that, for instance, Quine (e. g. 1960)—if we treat his view of stimulus meanings as an account of meaning at all—should be classified as a radical local holist, while thinkers like Sellars (e. g. 1974) or Ajdukiewicz (e. g. 1931, 1934) can be seen as global holists.

8 In other words, *for all meanings involved in a sentence uttered in a context, the meaning of a single expression depends on the meanings of all other expressions in that sentence.*

9 See Jackman (2015) for the view according to which Davidson's theory is of the kind that accepts GH and rejects LH.

within a given sentence, they are sensitive to each other's semantic properties. That last point clearly shows that endorsing LH does not entail accepting GH.

2.2 Contextualism[s]

Although the current debate concerning context-sensitivity has lasted for about two decades, one of the main issues within it is the lack of precise characterizations of the sides that are at odds. What definitely does not help is that there are several separate[10] aspects of context-sensitivity, and when you share someone's views concerning one of these aspects, it does not mean that you will necessarily agree with regard to another.[11] As my main purpose here is to analyze the relation between MH and Contextualism, I am going to set aside the intricate details involved in distinguishing between contextualists and anti-contextualists. Instead, I will extract the main features of Contextualism from several of its definitions given by the participants of the debate (see: Borg 2004, 2012; Cappelen & Lepore 2005; Recanati 2004, 2010; Stanley 2007). According to the picture that emerges from those characteristics, Contextualism is a view that can be condensed into the conjunction of the two following claims:

(**c1**) The intuitive propositional content of a sentence is sometimes determined by strong pragmatic effects.
(**c2**) The class of natural language context-sensitive expressions is significantly larger than the set of obviously indexical expressions (i.e. than 'Kaplan's set').

Let me start with an explanation of what *intuitive propositional content* is. As a matter of fact, 'propositional content' is a new term for the old-fashioned notion of *propositions*, the essential feature of which is that they are bearers of truth values and as such they are associated with truth-conditions (see: Kawczyński 2013). *Intuitive truth-conditions* are the truth-conditions corresponding to the proposition that are intuitively expressed by a given sentence.[12] For example,

10 Obviously, they are not entirely separate; however, they are distinct enough to cause the problems in question.

11 Joanna Odrowąż Sypniewska (2013 – 70) points out four different criteria according to which factions can be defined and shows that adopting different criteria sometimes ascribes the same philosopher to different sides of the debate. These criteria concern, respectively, the propositionality of semantic content, intuitive truth-conditions, the number of context-sensitive expressions and weak and strong pragmatic effects.

12 Intuitive truth-conditions are distinguished from something that may be called *formal* or *schematic* truth-conditions, which mirror the logical form of a given statement.

a sentence like 'John got sick and he went to hospital' is associated with the intuitive truth-conditions that *firstly John got sick and subsequently he (i.e. John) went to hospital* (as opposed to the *unintuitive*, formal truth-conditions determined by the logical definition of conjunction, which does not reflect the chronology of events) (see: Recanati 2004: 8–16; Stanley & Szabó 2000/2007: 25). Contextualists believe that intuitive truth-conditions can be determined—and according to the majority of them, that is what often happens—by so-called *strong* pragmatic effects that have been defined by Jeffrey King and Jason Stanley as follows: "A weak pragmatic effect on what is communicated by an utterance is a case in which context (including speaker intentions) determines interpretation of a lexical item in accord with the standard meaning of that lexical item. A strong pragmatic effect on what is communicated is a contextual effect on what is communicated that is not merely pragmatic in the weak sense" (King & Stanley 2005/2007: 140).

We may say that the processes governed by the standard meaning of a linguistic item are *semantics-driven*; then, the above distinction boils down to the rule that that every appeal to context which is not semantics-driven belongs to the class of strong pragmatic effects (see: Recanati 2010: 4). To mention but one of numerous examples of this kind of effect, imagine someone uttering 'Mary took the key and opened the door', which is understood as claiming that Mary took the key and opened the door *with that key*. What occurs in such a case is the so-called *free enrichment*[13] effect—the content of the utterance has been enriched, although it is not encoded in the semantics of the sentence that such an enrichment is compulsory.[14]

At this point the crucial question is of course how to distinguish what *is* from what *is not* semantics-driven. Even anti-contextualists, who aim to protect semantics from pragmatic influences, are not unanimous with respect to that question. Whereas minimalists like Borg (2004, 2012) or Herman Cappelen and Ernest Lepore (2006) believe that only explicit indexicals need context intervention, indexicalists like King, Stanley or Zoltan Szabó (see papers collected in: Stanley 2007) claim that indexicality is a feature of a considerably large class of expressions; although it is very often not recognisable on the (syntactic) surface, it is still present in their logical form. Contextualists do not give a clear answer either,

13 Recanati (2004: 25–26) distinguishes two further kinds of strong pragmatic effects. One is *loosening*, which is the converse of enrichment and basically consists in adopting looser

14 On the other hand, when someone utters e.g. 'I am hungry', it is encoded in the semantic constitution of the pronoun 'I' that you have to reach to context to interpret it (namely, you are supposed to figure out who the speaker is). Such a contextual enrichment, usually called *saturation*, is not 'free' and is semantics-driven.

and they often fall back on some *ad hoc* explanations based on linguistic intuitions (which usually end up with the conclusion: "it just has to be pragmatics—otherwise it would be incomprehensible!").

Setting controversial details aside, what is certain and essential about Contextualism is that intuitive truth-conditions are sometimes determined by strong pragmatic effects and—as entailed by (c2)—it is a common scenario. Thus, what may be considered Contextualism, in a nutshell, is the claim that *semantics and pragmatics interfere at every level of content-shaping*. According to minimalists and indexicalists, the output of semantics-driven processes is the input of strong pragmatic processes, while in Contextualism processes occurring at the very basic, even pre-propositional level, involve the interference of semantic and (strong) pragmatic factors.

As MH is a theory concerning *all* expressions, it seems that it is most closely related to a radical variety of Contextualism, in the *credo* of which the measured terms used in (c1) and (c2) are substituted by the resolute 'always' and 'all' (see: Recanati 2010: 18 – 19):

(**rc1**) The intuitive propositional content of a sentence is always determined by strong pragmatic effects.
(**rc2**) All natural language expressions are context-sensitive.

Although (rc1) and (rc2) are quite definite, it is still possible to distinguish two significantly different versions of Radical Contextualism (hereafter: 'RC') (see: Recanati 2004: 140 – 141) with regard to how strong the consequences of these claims are taken to be. One version, called the Wrong Format View, is the claim that although there are such things as literal meanings, they should be considered merely radicals or schemas of the meanings that genuinely count. In other words, literal meanings cannot enter intuitive propositions, because they do not have the proper *format* to do so—the literal meaning of a simple expression always has to be pragmatically adjusted to contribute to the semantic value of a complex expression. According to the other version of RC, called Eliminativism, there is nothing like literal meaning at all. The only input of a communicative act is the history of past uses of a given word (with the meanings that these words had on those occasions), while the output is the word in question with its meaning ascribed to this particular context of utterance. In other words, producing an expression-token in a given context does not begin with meaning of the expression-type, but rather with a review of the register of meanings of the relevant tokens from the past.

To have a clear vision of the different variants of Contextualism, in what follows I will use the term 'Moderate Contextualism' as denoting the theory that ac-

cepts the conjunction of (c1) and (c2) while at the same time rejecting (rc1) and (rc2). Such an account boils down to the conjunction of the two following claims:

(**mc1**) The intuitive propositional content of a sentence is sometimes-but-not-always determined by strong pragmatic effects.
(**mc2**) Many-but-not-all natural language expressions are context-sensitive.

3 Global Holism and Contextualism

Having sketched the theoretical landscape, I will move on to the main subject of my investigation—that is, the interactions between Holism[s] and Contextualism[s]. My methodology is fairly simple as it consists in examining whether two defined variants of MH *entail* or *are entailed by* various versions of Contextualism.

3.1 Global Holism and Eliminativism are not compatible

Let me start with a closer look at the relation between Global Holism and the three above-mentioned versions of Contextualism. There is not much to say about the strongest version of the latter theory, i.e. Eliminativism, since its only claim concerning literal meanings is that there are no literal meanings. Therefore, it is clear that no account of this kind can be reconciled with any theory of literal meanings such as GH (thus, in the rest of the section devoted to GH, when I talk about RC, rc1 and rc2, I treat them as interpreted in accordance with the Wrong Format View).

3.2 Global Holism and the Wrong Format View are compatible

The Wrong Format View (hereafter: 'WF') at least accepts the existence of literal meanings, so it is in principle possible to discover a connection between it and GH. Although the connection does indeed exist, it is not very impressive. Supporters of WF believe that each literal meaning has to be pragmatically formatted to become a semantic value of a given expression. There is no obstacle to their agreeing that those literal meanings have a holistic origin. On the other hand, a global holist can accept WF as the proper account of what happens

with literal meanings in contexts. Hence, it can be stated categorically that the theories in question do not exclude each other.

3.3 Global Holism and Moderate Contextualism are compatible

What practically differentiates Moderate Contextualism (hereafter: 'MC') from WF is the force of their statements: according to the former theory, there are *many-but-not-all* context-sensitive expressions that require strong pragmatic influences to become propositional, while, according to the latter, *all* expressions are of that kind. Thus, if GH is compatible with WF, which is apparently a stronger account than MC, it follows *a fortiori* that GH is also compatible with MC.

3.4 The Wrong Format View does not entail Global Holism

The next question is if there is any connection between GH and Contextualism that is stronger than mere compatibility. Let's start with figuring out if WF entails GH. I believe not. First, (rc1) does not lead to any thesis concerning the origin of literal meanings, which is the proper subject of GH. In other words, WF does not really concern the genesis of literal meanings and thus accepting that view does not commit one to any particular theory concerning meanings-before-contexts (which is precisely what GH is about). It does not matter if literal meanings are constituted holistically, atomistically or any other way. What an adherent of WF cares about is that the format they exhibit appears to be wrong, regardless of how they acquired it.[15]

Second, according to (rc2) *all* expressions are context-sensitive. Hence, it seems to me that to become a global holist, an adherent of WF would have to make some further assumptions; without making them, nothing forces her to accept GH. The first assumption she would have to endorse is something I call *the principle of essential changeability*. What I mean here stems from the following kind of reasoning: accepting WF is equivalent to believing that the meanings of *all* expressions should—and *can*—be adjusted by pragmatics; if you add to

15 Proponents of WF think that the literal meaning of e. g. the adjective 'green' cannot be its semantic contribution to the value of a complex expression including that word, because it has the wrong format; in such a case, it makes no difference if you think of meaning in a holistic manner i. e. as something determined by the totality of meanings of language, or if you are an atomist and believe that 'green' has its meaning regardless of the rest of language.

this the assumption that the fact that meanings reveal their flexibility at the level of particular context proves that they are *essentially changeable,* i.e. possible to change at every level,[16] then you cannot agree with the standpoint that all or some meanings are constant and stable i.e. (*essentially*) *unchangeable*. And to say that meanings are unchangeable would obviously go against holism. However, even making such a strong assumption is still not enough to entail becoming a global holist, since GH does not overlap with the principle of essential changeability for two reasons: (i) the principle of essential changeability is indeed an option for global holists but not a compulsory one, since they can maintain that meanings are changeable only at the level of their origin, while staying stable in contexts; (ii) GH is not the only theory that can endorse the principle—for instance, a molecularist can accept it as well and claim that every meaning in a language is indeed vulnerable to change however, it is not the case that each is dependent *on every other* meaning (but depends only on several closely related meanings). So now it seems not only that an adherent of WF is not obliged to accept GH, but also that in order to do that she has to do quite a lot. Namely, she must deny (i) and (ii), and that definitely requires some solid justification, as both these theses are quite intuitive. Hence, it turns out that endorsing GH is definitely not a must for her, and actually, rather than being a compulsory entailment, it requires her making a theoretical decision to endorse GH.

3.5 Moderate Contextualism does not entail Global Holism

Regarding (mc1), it does not lead to GH for precisely the same reasons for which (rc1) did not entail GH (see above).

Concerning (mc2), it can be said that if the prevailing context-sensitivity postulated in (rc2) does not commit one to accepting GH, then neither does the less widespread context-sensitivity of MC. What made WF allegedly similar to GH is that in both theories the vulnerability to changes is a feature of *all* meanings (and this vulnerability of meanings in general is responsible for causing the noted instability problems for both theories). In other words, if all meanings are vulnerable to contextual adjustments, as claimed in WF, there is a chance that they are also vulnerable to holistic changes at the global level. In MC, on the other hand, it is said that *many-but-not-all* words are context-sensi-

16 I do not want to judge how many levels there are and what exactly each of them could be. The two levels that are relevant to my discussion are: (1) literal meanings-formatting, which involves the perspective of language as a whole, and (2) literal meanings-formatting in contexts, as they eventually become components of propositions expressed by sentences in contexts.

tive, so in this case *many-but-not-all* expressions are potentially vulnerable to holistic changes, while GH is explicitly a theory concerning *all* expressions. Thus, for an adherent of MC, even endorsing the principle of essential changeability and denying (i) and (ii) is still not enough to entail being a holist. As a matter of fact, I cannot see any other way for a moderate contextualist to endorse GH than by by simply voluntarily doing so. In any case, there is definitely no entailment from MC to GH.

3.6 Global Holism does not entail Moderate Contextualism

Now we can move on to investigating whether GH entails some version of Contextualism. I am going to start with considering if GH entails MC, because if I can show that that is not the case, then *a fortiori* I will demonstrate that GH does not entail WF either.

The first question is whether GH entails the first thesis of MC, i.e. (mc1). I believe it does not for one very simple reason: a global holist is not obliged to endorse any particular solution concerning what the contribution of weak and strong pragmatic effects are to formatting meanings in contexts. That is so because the output of holistic processes at the level of a whole language is the possible input of weak or strong pragmatics in contexts. And nothing in the nature of those holistic processes favors this or that type of pragmatic effects as more suitable to affect literal meanings that are holistically constituted.[17]

Things are similar in the case of (mc2). Once again, a global holist is not committed to any particular view concerning the number of expressions that

17 For instance, according to global holists, the literal meaning of 'green' is defined by the relations in which it stands to every other meaning in English. Now, once its meaning has been constituted, a global holist may say that it is not her business how it will behave within contexts—in particular, whether it is sensitive to weak, strong or both types of pragmatic factors. It should be noticed, however, that a global holist can still offer some claims about the correlation between holistic constitution of meanings and their sensitivity to pragmatics in contexts. She can, for instance, explain why 'green' is sensitive to (let's say) only weak pragmatics by showing what the holistic relations that provide its meaning are with the format that it actually has. On the other hand, as I argue in Kawczyński (2017), a global holist can also adopt the position of Minimalism and claim that holistic processes, which have already constituted a given literal meaning, also make it insensitive to pragmatic effects—which shows even more clearly that GH does not entail Contextualism (as it is compatible with Minimalism, which is the polar opposite of Contextualism).

are context-sensitive as her theory is not a theory of meanings in contexts.[18] Yet, if for some reason we really desired to transform a global holist into a moderate contextualist, we would have to deal with essential changeability again, but this time on the holist's side. In this case, essential changeability is required to play the role of the bridge between the holist principle, which entails that all meanings are changeable, and the contextualist thesis that some meanings are context-sensitive. However, if accepting the rule concerning essential changeability—which would now take the form of something along the lines of: *that meanings reveal their flexibility at the level of language as a whole proves that at least some of them are essentially changeable i.e. able to change at every level, e.g. in contexts*—is not satisfactorily justified, it appears to be in fact nothing more than an *ad hoc* hypothesis advanced only in order to play the role of the link between GH and MC (without any reasons independent of that purpose). However, even if you take the principle at face value, accepting it is still not enough to find yourself in the contextualist framework. The principle does not determine that the changes in contexts are supposed to be strong rather than weak pragmatic effects—thus, to become a contextualist, a global holist would have to add to her theory the claim that every meaning is affected in context by *strong* pragmatics, which boils down to, once again, simply endorsing MC. There is no case, though, in which it can be said that the holistic view is a consequence that her original position leads to.

3.7 Global Holism does not entail the Wrong Format View

The reasons for denying that GH entails WF are almost identical to those that I have just offered to prove that GH does not lead to MC. The case of (rc1) does not differ from (mc1) with respect to being possibly entailed by GH. Regarding (rc2), the only difference is that the essential changeability principle would take a stronger form than in the case of (mc2), namely: *that meanings reveal their flexibility at the level of language as a whole proves that all of them are essentially changeable i.e. able to change at every level, e.g. in contexts*. And since this formulation is stronger it is even more controversial and thus—so to speak—*more ad hoc*. Still, such a principle does not commit anyone to believing that the

18 However, just as in the case of (mc1), holists can have something to say about that, and again, there is nothing preventing them from simply endorsing the minimalist account, according to which a strictly limited group of expressions is context-sensitive (see Kawczyński (2017) for more details).

changes within contexts are always strongly pragmatic, and so it does not lead to Contextualism.[19]

4 Local Holism and Contextualism

The general conclusion of the previous section was that there is no strong correlation between Global Holism and any of the discussed variants of Contextualism. It turns out that the instability that is allegedly responsible for why GH and Contextualism are often tarred with the same brush manifests in GH at a different level than in Contextualism. Unlike GH, LH actually crosses paths with Contextualism as both theories concern the behavior of meanings within sentences uttered in contexts. Moreover, none of the three discussed versions of Contextualism can be excluded as incompatible with LH.

4.1 Local Holism is compatible with Eliminativism, the Wrong Format View and Moderate Contextualism

I put the three above relations in one pot, because I believe it to be rather evident that LH is compatible with all these variants of Contextualism. On the one hand, a contextualist can claim that contextual processes responsible for *construing* all meanings (Eliminativism), *formatting all* meanings (WF) or *formatting many-but-not-all* meanings (MC), are holistic in nature. A holist, on the other hand, is able to choose between these three views of context-sensitivity. The only doubts that can be raised regard the third option, namely MC, since LH is about *all*, while MC is about *many-but-not-all* meanings in contexts. However, reconciling LH and MC is still possible when not all holistic processes involved in a context are thought to be based on strong pragmatics. In other words, a holist is able to take the pragmatic effects that MC is about[20] to be a proper subset of all holistic processes that shape meanings in a context. Roughly speaking, some meanings appearing in a context are said to be affected by strong or weak pragmatic processes, while others are affected by holistic factors that are not associated with context-sensitivity. What those non-contextual-holistic processes are supposed to be is an interesting question to which I am going to come back to later (see section 4.5

19 Analogously to the case of MC, one could also show that accepting the minimalist position is a realistic option for a global holist, which further undermines the claim that a proponent of GH is obliged to accept WF.

20 For the moment, I ignore the question of whether they are strong or weak.

below). For now, I just want to point out that there exists such a theoretical possibility and that it makes LH and MC compatible.

4.2 Neither Eliminativism, the Wrong Format View, nor Moderate Contextualism entail Local Holism

I suppose that this point is even less problematic than the previous one. As I said in the preceding paragraph, a contextualist (of any kind) is able to adopt the holistic method to characterize the contextual processes that construe/format meanings. I would only add that she simply does not have to do that, and that she can choose an atomistic or molecularist approach just as well.[21]

4.3 Local Holism does not entail Eliminativism

Symmetrically to what I have said about GH (namely, that it does not consider meanings-in-context—see 2.6 above), let me point out now that LH as a theory of meanings-in-context does not lead to any particular views about literal meanings, especially not to the eliminativist view according to which there are no literal meanings. As shown before,[22] LH is compatible with GH as well as with WF and MC—all of which are the theories accepting literal meanings, so it is clear that LH cannot entail Eliminativism (although, as noticed in 3.1 it is possible to reconcile them).

21 A contextualist, let's say an adherent of WF, may think that in the case of my uttering e. g. 'I am green', the meanings of each word in the sentence require some pragmatic influence to obtain the proper format and become a part of a proposition. However, there is nothing that forces her to believe that formatting the meaning of 'I' depends on the formatting of 'am' and 'green'. She can endorse the view that each of these words is influenced by the context, but they do not influence each other—and thus she becomes an atomist. Or, she could say that for instance 'I' and 'green' are interdependent, but neither of them depends on 'am'—and thereby become a molecularist. The same would happen in case of Eliminativism (and meaning construal instead of format) and MC (and formatting not all, but some meanings).

22 See sections 2.1.1 and 3.6 above.

4.4 Local Holism does not entail the Wrong Format View

In Kawczyński (2017) I offer an elaborated argument in favor of the thesis that LH is compatible with Minimalism.[23] To put it briefly, I present a version of LH in which the holistic processes that affect all expressions in contexts are of two kinds: weakly pragmatic or purely semantic. The latter processes play the role of *blockers* of pragmatic processes, i.e. they prevent literal meanings from being influenced by any pragmatic effects. Such a variant of LH is incompatible with Contextualism (of any kind) as it excludes strong pragmatics from the process of formatting propositions. I defined the purely semantic processes in particular to demonstrate that LH is not doomed to be inconsistent with the minimalist thesis according to which *some-but-not-many* expressions are sensitive to the weak pragmatic effects. To show that LH does not entail WF, even less is necessary. Namely, it is sufficient to notice that a local holist can expand her theory with the claim that all holistic processes are semantics-driven i.e. are subject to weak pragmatic effects. That move amounts to a rejection of (rc1) (as interpreted in WF), and in turn, confirms that LH does not lead inevitably to WF.

4.5 Local Holism does not entail Moderate Contextualism

What has been said in the preceding point also applies to MC: LH involving only purely semantic or weakly pragmatic processes entails the rejection of (mc1) as well. However, the interpretation of LH presented above is not the only one that is inconsistent with MC. A less controversial version of LH, which does not involve purely semantic processes and at the same time allows both weak and strong pragmatic processes, cannot lead to MC either. That is so because according to such a holism *all*[24] expressions are context-sensitive which clearly runs contrary to (mc2). Moreover, a version of LH in which it is stipulated that all holistic processes are strongly pragmatic also excludes MC because such a version entails the rejection of both (mc1) and (mc2).[25]

---

23 And in principle, if a theory is compatible with Minimalism, *ceteris paribus* it does not inevitably lead to Contextualism.

24 Because in LH, by definition, *all* meanings go through holistic processes, so if it is stipulated that holistic processes are weak pragmatic effects and strong pragmatic effects, then it follows that all meanings are pragmatically affected.

25 If all meanings are affected by strong pragmatics, then it goes against the 'sometimes-but-not-always' part of (mc1) and flagrantly violates the 'many-but-not-all' condition involved in (mc2).

5 Conclusion

The main results of the above investigation can be summarized in the four following claims:

1. Global Holism is compatible with all versions of Contextualism, except Eliminativism.
2. Global Holism neither entails nor is entailed by any version of Contextualism.
3. Local Holism is compatible with all versions of Contextualism.
4. Local Holism neither entails nor is entailed by any version of Contextualism.

The general result is thus as follows: when it comes to the issue of context-sensitivity, a meaning holist, either global or local, has a variety of options to choose between, and *vice versa* a contextualist of this or that kind can choose to accept either Local or Global Holism, to accept both, or to avoid both. In every case, however, committing to another view requires one to expand one's original theory with some further assumptions. This proves that the reasons to endorse a version of MH or Contextualism are mutually independent and that being, for instance, a holist and at the same time being attracted to contextualist ideas, one is supposed to make some effort to become a holistic contextualist (or a contextualist holist). Yet, if a holist does not have that type of contextualist inclination, she is able to reject Contextualism (and become a minimalist, which is possible as I have already mentioned). It works analogously the other way around—if you are a contextualist, you are not obliged to any version of holism, but it is still possible to endorse MH provided that you expand your theory with some additional adjustments.

One may complain that what I have argued is in fact quite a weak thesis according to which Contextualisms and Holisms are *logically* independent, while *in practice,* if someone commits themselves to one of them, one typically commits also to the other one. I believe such complaints arise directly from the prejudices that have become a kind of methodological dogma of taking MH and Contextualism to be close relatives. In this paper, I have tried not only to justify the claim that no entailment-relation holds between the theories discussed, but I have also sketched possible ways[26] of modifying MH or Contextualism that make them theoretically closer or more distant.

26 E.g. endorsing the principle of essential changeability or accepting some claims about the nature of holistic processes.

All in all, my answer to the question involved in the title of the paper is the following: neither friends, nor foes, but something else. If I were to continue this personification-metaphor, I would say that Meaning Holism and Contextualism are distant relatives, living far away from each other, having grown up in significantly different social and natural environments, but having similar interests, who could possibly—in favorable circumstances and with a dose of goodwill from both—become friends. Or rivals, if the circumstances and intentions proved to be less than amicable.

References

Ajdukiewicz, Kazimierz. 1931. "O znaczeniu wyrażeń". *In Język i poznanie*. (Vol. 1, pp. 102–136). Warsaw: PWN. 1985.

Ajdukiewicz, Kazimierz. 1934. "Language and Meaning". In Jerzy Giedymin (ed.) *The Scientific World-Perspective and Other Essays, 1931–1963*. Dordrecht: Springer Netherlands. 35–66.

Bilgrami, Akeel. 1998. "Why Holism is Harmless and Necessary". *Philosophical Perspectives 12*(S12): 105–126.

Block, Ned. 1994. "An Argument for Holism". *Proceedings of the Aristotelian Society* 95(n/a): 151–170.

Borg, Emma. 2004. *Minimal Semantics*. (Vol. 116). Oxford: Oxford University Press.

Borg, Emma. 2012. *Pursuing Meaning*. Oxford: Oxford University Press.

Brandom, Robert. 2000. *Articulating Reasons: An Introduction to Inferentialism*. (Vol. 110): Cambridge (MA): Harvard University Press.

Brandom, Robert. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. (Vol. 183). Cambridge (MA): Harvard University Press.

Cappelen, Herman and Lepore Ernest. 2005. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. (Vol. 117), John Wiley & Sons

Cappelen, Herman and Lepore Ernest. 2006. "Insensitive Semantics". *Philosophy and Phenomenological Research 73*(2). 443–450.

Churchland, Paul Montgomery. 1993. "State-Space Semantics and Meaning Holism". *Philosophy and Phenomenological Research 53*(3), 667–672.

Devitt, Michael. 1996. *Coming to Our Senses: A Naturalistic Program for Semantic Localism*. Cambridge: Cambridge University Press.

Field, Hartry. 1977. "Logic, Meaning, and Conceptual role". *Journal of Philosophy 74*(July), 379–409.

Fodor, Jerry, A. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.

Fodor, Jerry, A. and Lepore Ernest. 1992. *Holism: A Shopper's Guide*. (Vol. 43): Blackwell.

Fodor, Jerry A. and Lepore Ernest. 2002. *The Compositionality Papers*. Oxford University Press.

Harman, Graham. 1993. “Meaning Holism Defended”. *Grazer Philosophische Studien 46*, 163 – 171.
Jackman, Henry. 1999. “Moderate Holism and the Instability Thesis”. *American Philosophical Quarterly 36*(4), 361 – 369.
Jackman, Henry. 2015. “Externalism, Metasemantic Contextualism, and Self-knowledge”. In Sanford Goldberg (Ed.), *Externalism, Self-Knowledge, and Skepticism: New Essays*. New York: Oxford University Press. 228 – 247.
Jackman, Henry. 2017. “Meaning Holism”. In Edward Zalta (ed.) The Stanford Encyclopedia of Philosophy: Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2017/entries/meaning-holism/
Kawczyński, Filip. 2013. “Czy filozofia języka potrzebuje pojęcia sądu logicznego?”. In Piotr Stalmaszczyk (ed.) *Współczesna filozofia języka. Inspiracje i kierunki rozwoju*. Łódź: Primum Verbum. 111 – 135.
Kawczyński, Filip. 2017. “Is Meaning Holism Compatible with Semantic Minimalism?”. *Studia Semiotyczne XXXI*(2), 53 – 75.
King, Jeffrey and Stanley Jason. 2005/2007. “Semantics, Pragmatics, and the Role of Semantic Content”. In *Stanley 2007*. Oxford University Press. 133 – 181.
Kölbel, Max. 2004. “Faultless Disagreement”. Proceedings of the Aristotelian Society 104(1), 53 – 73.
Lormand, Eric. 1996. “How to be a Meaning Holist”. *Journal of Philosophy 93*(2), 51 – 73.
Odrowąż-Sypniewska, Joanna. 2013. *Kontekstualizm i wyrażenia nieostre*. Warszawa: Semper.
Pagin, Peter. 1997. “Is Compositionality Compatible with Holism?”. *Mind and Language 12*(1), 11 – 33.
Pagin, Peter. 2006. “Meaning Holism”. In Ernest Lepore and Barry. C. Smith (eds.), *The Oxford Handbook of Philosophy of Language*. Oxford University Press. 213 – 232.
Peacocke, Christopher. 1997. “Holism”. In Bob Hale and Crispin Wright (eds.) *A Companion to the Philosophy of Language*. Wiley-Blackwell. 227 – 247.
Penco, Carlo. 2001. “Local Holism”. In Paolo Bouquet (ed.) *Lecture Notes in Artificial Intelligence*. Kluwer Academic Publishers. 290 – 303.
Quine, Willard Van Orman. 1960. *Word and Object*. MIT Press.
Quine, Willard Van Orman. 1951. “Two Dogmas of Empiricism”. *Philosophical Review 60*(1), 20 – 43.
Recanati, François. 2004. *Literal Meaning*. Cambridge University Press.
Recanati, François. 2010. *Truth-Conditional Pragmatics*. Oxford University Press.
Rovane, Carol. 2013. “The Larger Philosophical Significance of Holism”. In Ernest LePore and Kirk Ludwig (eds.) *A Companion to Donald Davidson*. Wiley Blackwell. 395 – 409.
Sellars, Wilfrid. 1974. “Meaning as Functional Classification”. *Synthese 27*(3 – 4). 417 – 437.
Stanley, Jason. 2000. “Context and Logical Form”. *Linguistics and Philosophy 23*(4). 391 – 434.
Stanley, Jason. 2002. “Making it articulated”. *Mind and Language 17*(1&2). 149 – 168.
Stanley, Jason. 2005. “Semantics in Context”. In Gerhard Preyer and Georg Peter (eds.). *Contextualism in Philosophy: Knowledge, Meaning, and Truth*. Oxford University Press. 221 – 254.
Stanley, Jason. 2007. *Language in Context: Selected Essays*. Oxford University Press.
Stanley Jason and Gendler Szabó Zoltan. 2000/2007. “On Quantifier Domain Restriction”. In (*Stanley 2007)* (Vol. 15). Oxford University Press. 69 – 110.

Marián Zouhar

On the Nature of Non-Doxastic Disagreement about Taste

Abstract: Disagreements about matters of personal taste are a rather resilient kind of phenomenon that calls for an unorthodox solution. Hybrid taste-expressivism seems to provide such a solution because it explains taste disagreements in non-doxastic terms. According to it, utterances of taste sentences play a dual role—on the one hand, they are used to express propositions in which objects are ascribed taste properties and, on the other hand, they are used to manifest evaluations of the objects contained in the propositions. The non-doxastic nature of taste disagreements is derived from the latter. My aim is to cast some doubt on the prospects of the hybrid taste-expressivist account. Although I admit that this explanation is on the right track in treating taste disagreements as non-doxastic, hybrid taste expressivism does not seem to be general enough. There are some instances of taste disagreement that are not amenable to this kind of explanation. I provide an alternative explanation that retains the non-doxastic nature of taste disagreements, but does not suffer from its limited explanatory power.

1 Introduction

There is an obvious difference between *manifesting* an evaluative attitude toward something and *saying that* one bears an evaluative attitude toward something.[1] Intuitively, if the speaker says 'Yum!' upon trying a delicious chocolate cake, she manifests her evaluative attitude toward the cake without saying that she bears the evaluative attitude in question. On the other hand, if the speaker utters 'This chocolate cake tastes good to me', she dispassionately *describes* what her evaluative attitude toward the cake is. It seems plausible that in describing her attitude toward the cake she is not manifesting the attitude; rather, she is merely dispassionately saying that she bears the attitude in question.[2] There are various

1 This point is forcefully made in Buekens (2011). In fact, it is a starting point of Buekens' criticism of both contextualist and relativist approaches to disagreements about matters of personal taste.

2 The same kind of difference occurs between exclaiming 'Ouch!' and saying 'I am in pain', for example.

Marián Zouhar, Comenius University in Bratislava, Faculty of Arts (marian.zouhar@uniba.sk)

https://doi.org/10.1515/9783110702286-004

systematic differences between these two kinds of utterances and their communicative effects; the difference that is most important for our present purposes consists in that the latter utterance contains an act of expressing a proposition while the former contains no such act.

Based on this, we may distinguish between *expressively* uttered taste locutions and *descriptively* uttered taste sentences. An expressively uttered taste phrase serves to manifest one's evaluative attitude toward something, and a descriptively uttered taste sentence serves to inform one's audience that one bears an evaluative attitude toward something. The interjections like 'Yum!' and 'Yuck!' illustrate the former while the explicitly relational taste sentences like 'This chocolate cake tastes good to me' and 'I dislike the taste of this chocolate cake,' which contain a direct reference to the speaker, illustrate the latter.

It is sometimes argued that some locutions can be simultaneously used in both ways. It seems that taste sentences that are not explicitly relational are a case in point;[3] a typical example is 'This chocolate cake is delicious'. It is claimed that when one utters 'This chocolate cake is delicious' one both *describes* what one's evaluative attitude toward the taste of the cake is and *manifests* the attitude in question—the sentence is used both descriptively and expressively.[4] Alternatively, it can be said that an utterance of 'This chocolate cake is delicious' has two kinds of content—it both expresses a proposition in which the speaker's attitude toward something is described and expresses a non-propositional content that consists in manifesting the speaker's attitude in question. Using the standard terminology, the utterance simultaneously has both *doxastic* and *non-doxastic* content. The doxastic content consists in expressing a taste proposition and the non-doxastic content consists in expressing an evaluative attitude toward the thing that is contained in the proposition.

This kind of approach to taste utterances and their semantics is currently in vogue. It is usually developed under the heading of *hybrid expressivism* or, more

3 The difference between explicitly relational taste sentences and taste sentences that are not explicitly relational is merely provisional. The illustrations given in the main text should give a clue as to how the difference is to be understood. Explicitly relational taste sentences contain references to both agents and objects of taste; they further contain binary predicates that express relations in which the agents stand to the objects. Taste sentences that are not explicitly relational are, as a rule, of the subject-predicate form; they contain references to objects of taste and unary predicates that express taste properties. Later on, I will distinguish between taste sentences (understood as not being explicitly relational) and somewhat narrowly conceived attitudinal sentences. The latter difference is, however, supposed to be independent of the former difference.
4 Again, Buekens (2011) presents this point convincingly.

appropriately, *hybrid taste-expressivism*.[5] Clearly, hybrid taste-expressivism is attractive for several reasons. In particular, it is sometimes presented as a suitable kind of approach because it can plausibly explain various instances of disagreements about matters of personal taste. Nevertheless, my aim in this paper is to cast some doubt on the prospects of hybrid taste-expressivist accounts of taste disagreements. It seems to me that there are some specific instances of taste disagreement that cannot be explained in a satisfactory manner by hybrid taste-expressivist accounts. Despite this, however, I take it that this account of taste disagreements is essentially on the right track. Based on this, I elaborate an alternative explanation that is close enough to hybrid taste-expressivism, but does not suffer from its somewhat limited explanatory power.

The structure of the paper is as follows. I start with presenting the core of hybrid taste-expressivist accounts of taste disagreements (Section 2). I then develop two kinds of cases that do not seem to be amenable to a hybrid taste-expressivist explanation (Section 3). Section 4 contains an outline of an alternative explanation that is close enough to hybrid taste-expressivism but has much broader explanatory power. This is illustrated by its application to the cases discussed in Section 3. Section 5 concludes the paper by giving a summary of the main results achieved.

2 Hybrid Taste-Expressivism and Disagreement about Taste

It is widely admitted among philosophers that when one utters 'This chocolate cake is delicious' one says something about one's own subjective view on a particular cake's taste rather than something about objective facts concerning the cake.[6] Based on this, the point of certain kinds of exchanges may appear to be rather abstruse. For example, it might seem incomprehensible why Bert presents

5 Theories that can be classified as (closely resembling) hybrid taste-expressivist accounts can be found, most notably, in Buekens (2011), Clapp (2015), Gutzmann (2016), Huvenes (2012), (2014) and Marques & García-Carpintero (2014). Marques (2016) develops a hybrid expressivist approach to aesthetic predicates. Hybrid expressivism is traditionally associated with theories that concern moral discourse; some recent proposals can be found in Boisvert (2008), (2014), Copp (2001), (2014) and Eriksson (2014).

6 For a criticism of objectivism see, for example, MacFarlane (2014: 2–7). Hirvonen (2016) suggests that our 'folk' theory of taste is objectivist, and proposes an objectivist semantics for taste predicates. At the same time, however, she claims that metaphysical objectivism about taste is incorrect. Recently, Wyatt (2018) defended an absolutist semantics for taste utterances.

himself as disagreeing with Alice in the following dialogue (even though both of them are talking about the same chocolate cake):

(1) Alice: This chocolate cake is delicious.
Bert: I disagree. This chocolate cake is not delicious.

If Alice says something about her own subjective view on the cake's taste and Bert responds by saying something about his own subjective view on the same cake's taste, why are they to be portrayed as disagreeing with one another? What is the point of Bert's utterance of 'I disagree'? Clearly, Bert's bearing a depreciative attitude toward the cake's taste is actually fully compatible with Alice's bearing an appreciative attitude toward it.

What is really puzzling is that we do seem to have a well-entrenched pre-theoretical intuition that Bert disagrees with Alice in (1). This fact calls for an explanation. If Alice is portrayed as saying something about her own tastes concerning a particular cake and Bert is portrayed as saying something about his own respective tastes concerning the same cake, the claim that Bert disagrees with Alice does not seem to be warranted. Yet, our intuition says that he does disagree with her. Bert somehow rejects what Alice says, although he hardly intends to reject that Alice's view on the cake's taste is appreciative. This is a perplexing situation; in fact, its perplexity is what makes taste disagreements a rather resilient kind of phenomenon that seems to call for an unorthodox explanation.[7]

Now, hybrid taste-expressivism seems to be precisely such an unorthodox suggestion. It provides a neat explanation that is able to dissolve the above perplexity. Hybrid taste-expressivism admits that Alice's and Bert's utterances express propositions that concern their own idiosyncratic tastes; since these propositions are fully compatible, there is no disagreement between Alice and Bert at this level—it is granted that Bert does not reject the proposition expressed by Ali-

7 Several types of approaches were proposed in the literature, and the list of works on taste disagreement and related issues is very extensive. I do not wish to discuss the various types of approaches in this paper. I just wish to point out that a considerable part of the literature is devoted to what is called *faultless disagreement*. Some philosophers allow that taste disagreements can be—or appear to be—faultless (e. g. Barker 2010, Buekens 2011, Egan 2014, Kölbel 2003, Lasersohn 2005, López de Sa 2008, Palmira 2015, Schaffer 2011). On the other hand, other philosophers reject the very idea of faultless disagreement (e. g. Cappelen & Hawthorne 2009: Ch. 4, Iacona 2008, Rosenkranz 2008, Smith 2010, Stojanovic 2007, 2011, Zouhar 2014). Recanati (2007, 88 – 94) argues that faultless disagreements may arise only in specific situations, but in ordinary situations disagreements are not faultless. I take it that taste disagreements are a very problematic kind of phenomenon independently of whether they are supposed to be faultless or not. That is why I ignore the question of their presumed faultlessness.

ce's utterance. In other words, Bert cannot be correctly portrayed as disagreeing with Alice in the doxastic sense of disagreement, and hybrid taste-expressivism subscribes to this position.

At the same time, however, hybrid taste-expressivism admits that, apart from having their respective doxastic contents, the taste sentences uttered in (1) have a non-doxastic kind of content. According to this approach, a single utterance of taste sentence has both a descriptive dimension and an expressive dimension. Thus, Alice and Bert use their respective lines both descriptively and expressively, in which case they succeed in manifesting their evaluative attitudes toward a particular cake. Clearly, the two attitudes are non-cotenable because it is impossible for someone to entertain both attitudes without considerably changing one's mind.[8] Thus, if Alice manifests her appreciative attitude toward the cake and Bert responds by manifesting his unfavorable attitude toward the same cake that is non-cotenable with Alice's attitude, Bert makes clear his rejection of Alice's attitude. It means that he disagrees with her at the non-doxastic level. As a result, hybrid taste-expressivism explains the kind of disagreement contained in the exchanges like (1) as non-doxastic because one party rejects the other party's evaluative attitude toward something, although the former does not intend to reject the latter's evaluative attitude in question.

This hybrid taste-expressivist account is based on a close alliance between two things, namely the idea that, apart from expressing propositions, taste sentences are used to express evaluative attitudes and the idea that taste disagreements are portrayed as non-doxastic. It might thus seem that the non-doxastic treatment of taste disagreements is made possible by the fact that taste sentences can be used expressively. In particular, the very fact that Bert manifests his evaluative attitude toward the cake in question in connection with the fact that his attitude is non-cotenable with the attitude Alice shows toward the same cake constitutes Bert's disagreement with Alice.

The close alliance between these two things seems to be crucial for hybrid taste-expressivism. It is tacitly assumed that all instances of non-doxastic taste disagreements are related to the expressive kind of use by which the speakers are claimed to manifest their respective evaluative attitudes toward something. Now I am going to show that this is not the case—there seem to be instances of non-doxastic taste disagreement in which the speakers do not use their respective taste sentences expressively in order to manifest their evaluative attitudes toward something. If this line of reasoning is correct, hybrid taste-expressivism is in trouble because there are instances of non-doxastic taste

8 On the notion of non-cotenability see, in particular, MacFarlane (2014: 121–123).

disagreements that are not amenable to the hybrid expressivist type of explanation. Nevertheless, I do not intend to undermine the very idea that taste disagreements are best explained in non-doxastic terms. The explanation to be proposed in Section 4 indeed retains this idea, but discards the hybrid taste-expressivist view that the non-doxastic nature of taste disagreements has to be based on invoking the expressive dimension of taste sentences.

3 Non-Doxastic Disagreements about Taste without Expressive Use

In this section, I try to divorce the idea that taste disagreements are best explained as non-doxastic from the idea that speakers should use their taste sentences expressively in order to disagree with their interlocutors in a non-doxastic manner. My aim is to preserve the former and reject the latter. I hasten to add, however, that the following considerations are not intended to undermine the hybrid taste-expressivist approach in general; rather, they are focused on one particular type of hybrid taste-expressivism, namely the one instantiated mainly by Buekens's account developed in his (2011), in which manifesting evaluative attitudes is deeply intertwined with the expressive use of taste sentences.[9]

There are at least two kinds of situations in which the speakers can be understood as expressing non-doxastic disagreements with their interlocutors, but do not use their taste sentences expressively. First, this may happen when speakers adopt an exocentric perspective, i.e. when they are talking as if taking up someone else's standpoint.[10] It seems to me that speakers may use taste sentences to express non-doxastic disagreements even when taking up the exocentric perspective.[11] At the same time, it is admitted that, when adopting an exocentric perspective, the speakers do not use taste sentences expressively. Second, the

9 For instance, it seems to me that Gutzmann's hybrid taste-expressivist account developed in his (2016) that is based on the general use-conditional semantics from Gutzmann (2015) is not threatened by the examples given below in the main text. Nevertheless, there are some other problems with the other hybrid taste-expressivist accounts; I leave their discussion to some other time.

10 On the difference between exocentric perspective and autocentric perspective, see Lasersohn (2005: 670).

11 This is explicitly rejected by Buekens; more precisely, he rejects that, when adopting an exocentric perspective, speakers may use taste sentences expressively and, based on this, he rejects that they may disagree with one another in the non-doxastic sense. See, in particular, Buekens (2011: 642 and 649).

same effect may be achieved when speakers adopt an autocentric perspective—i. e. when they are taking up their own standpoint—but something makes it impossible that their utterances receive an expressive kind of use. It seems to me that this happens when taste sentences are embedded within larger linguistic constructions of certain types. Nevertheless, speakers may express non-doxastic disagreement with their interlocutors even in this kind of situation.

If these lines of reasoning are correct, a different explanation is required that would retain the idea that taste disagreements are non-doxastic, but reject the idea that expressive use is essential to expressing such disagreements.

3.1 Exocentric Perspective and Non-Doxastic Disagreement

I will start with the first kind of situation. If the speakers adopt an exocentric perspective, they are speaking as if adopting someone else's point of view. Their utterances are to be evaluated as true or false relative to a third party's standards of taste or, alternatively, their utterances express propositions that contain the taste properties that are determined relative to a third party's standards of taste.[12] If one speaker utters 'That chocolate cake is delicious' on the basis of watching a third party eating the cake and making her pleasure obvious, and another speaker responds with 'No. That chocolate cake is not delicious' on the basis of watching the same scene but misinterpreting the signs of pleasure as signs of disgust, both speakers adopt the exocentric perspective that belongs to the third party. Clearly, they disagree. Their disagreement, however, is doxastic because the second speaker disagrees with the proposition expressed by the first speaker's utterance. Moreover, it is easy to see that they are not using their sentences to manifest their own or someone else's evaluative attitudes toward the cake—their utterances are descriptive rather than expressive. It is

12 The first half of the sentence to which this footnote is appended reminds of the relativistic approach to taste utterances while the second half captures the contextualist approach. The best developed versions of relativism can be found in Kölbel (2002), Lasersohn (2005), MacFarlane (2014), Richard (2008), and Wright (1992). Contextualism concerning taste predicates is adopted, for example, in Glanzberg (2007), Huvenes (2012), Marques & García-Carpintero (2014), Schaffer (2011), and Silk (2016). I prefer the contextualist treatment, as will become obvious in what follows. The contextualist semantics for taste utterances can be nicely combined with hybrid taste-expressivism, as the controversy between relativism and contextualism illustrates; I leave this controversy aside in this paper, but see Zouhar (2018) for a discussion. See also Baker (2012), Barker (2013), Egan (2014), López de Sa (2008), Plunkett & Sundell (2013), Stojanovic (2007), Sundell (2011), and Zeman (2016) and (2017) for further discussions and proposals.

tempting to extend this observation to all instances of taste disagreement that are made relative to exocentric perspectives.

If this extension is correct then the connection between the two ideas mentioned in the preface to this section may seem to hold. However, I am going to show that the extension actually fails. It is rather easy to devise situations in which (*i*) speakers adopt an exocentric perspective, (*ii*) they are intuitively supposed to disagree with one another, and (*iii*) their disagreement is best explained as non-doxastic (rather than doxastic). These are situations in which the speakers do not use their sentences expressively and, yet, their disagreement is non-doxastic. This suffices to break the supposed connection between the two ideas. Let me elaborate on an example that illustrates this point.

Suppose that Alice and Bert are trying to discern delicious meals from those that are not delicious by observing manifest responses of two subjects to eating various kinds of meals. Alice observes Cynthia's behavior and Bert Dave's. Alice observes that upon trying a particular chocolate cake—call it 'Choco'—Cynthia makes her pleasure obvious, while Bert finds out that upon trying from the same cake, Dave responds with clear signs of disgust. Based on these observations, the following exchange takes place, where Alice comments on the results based on Cynthia's reactions and Bert comments on the results based on Dave's reactions:

(2) Alice: Choco is delicious.
Bert: No. Choco is not delicious.

Both Alice and Bert adopt an exocentric perspective. Alice says that the cake is delicious because it seems to taste good to Cynthia, and Bert says that the cake is not delicious because it does not seem to taste good to Dave.

Obviously, this exchange contains disagreement. Despite adopting exocentric perspectives, the disagreement between the speakers is not doxastic. Bert does not object to Alice's proposition in which Choco is claimed to be delicious according to Cynthia's standards of taste. In other words, Bert does not charge Alice for mischaracterizing Cynthia's tastes concerning Choco. Their disagreement is rather non-doxastic. Of course, Alice cannot be interpreted as manifesting her own evaluative attitude toward the cake and Bert cannot be interpreted as manifesting his own evaluative attitude that would be non-cotenable with Alice's. Adopting Cynthia's perspective, Alice presents Cynthia's non-doxastic attitude while Bert, adopting Dave's perspective, presents Dave's non-doxastic attitude that is non-cotenable with Cynthia's.

Importantly, neither Alice nor Bert are using their sentences expressively. Their utterances are merely descriptive. Despite this fact, they succeed in manifest-

ing their respective subjects' evaluative attitudes. This case makes it clear that expressing non-doxastic taste disagreement can be detached from using taste sentences expressively. Based on this, it is obvious that hybrid taste-expressivism cannot explain this kind of disagreement along the same lines as it explains other kinds. Thus, hybrid taste-expressivism does not seem to be general enough.

3.2 Embedded Utterances and Non-Doxastic Disagreement

Another problematic kind of situation obtains when (*i*) speakers utter taste sentences that are embedded within larger linguistic constructions, (*ii*) they are intuitively supposed to disagree with one another, and (*iii*) their disagreement is best explained as non-doxastic. Since in such situations the speakers do not use their taste sentences expressively, the supposed connection between the two ideas outlined in the preface to this section breaks down once more.

Here is a situation in which all three points seem to obtain. Assume that Alice's tastes are very similar to Cynthia's and that Bert's tastes are radically different from Cynthia's. It means, roughly, that if Cynthia finds some meal tasty, it is usually tasty for Alice but not for Bert. Now the following exchange occurs:

(3) Alice: If Cynthia finds Choco delicious, then it is delicious.
Bert: No. If Cynthia finds Choco delicious, then it is not delicious.

Based on her previous experiences according to which her tastes are rather close to Cynthia's, Alice describes Choco as delicious provided it is tasty for Cynthia, and she does so relative to her own standards of taste. Similarly, based on his previous experiences according to which his tastes are rather different from Cynthia's, Bert describes the same cake as not delicious provided it is tasty for Cynthia, and he does so relative to his own standards of taste. Intuitively, Bert disagrees with Alice. Nevertheless, Bert cannot be said to reject the proposition expressed by Alice's utterance of 'It is delicious'. It means that his disagreement is not doxastic.

Clearly, Bert adopts a different evaluative attitude toward the cake as Alice, and, based on this, his disagreement with Alice is best explained as non-doxastic. At the same time, Alice's and Bert's utterances of 'It is delicious' and 'It is not delicious', respectively, are used descriptively rather than expressively. In fact, neither Alice nor Bert have tried Choco, and thus they cannot sincerely manifest their attitudes toward the cake. Despite this fact, they do express their respective evaluative attitudes, although they do so in a conditional way.

This kind of situation shows that manifesting one's own evaluative attitudes toward something does not require that one uses one's taste sentence expressively. As a result, this kind of situation does not seem to be amenable to the hybrid taste-expressivist treatment according to which an expressive use of a taste sentence is a prerequisite of expressing one's non-doxastic disagreement with someone else's evaluative attitude toward something. Thus, what is required is an alternative explanation that would both retain the non-doxastic nature of taste disagreements and avoid invoking the expressive kind of use of taste sentences. The following section contains an outline of such an explanation.

4 A Requisite-Based Account of Non-Doxastic Disagreement about Taste

The explanation to be developed in this section coincides with hybrid taste-expressivism in certain vital respects. First, it takes disagreements about taste to be non-doxastic. Second, it claims that by uttering taste sentences the speakers express propositions in which certain objects are ascribed certain properties. Third, it admits that by uttering taste sentences the speakers make obvious their evaluative attitudes toward something. These resemblances notwithstanding, there is one crucial difference—the explanation developed below does not assume that the speakers make their evaluative attitudes obvious by using their taste sentences in the expressive way. A different mechanism for manifesting the evaluative attitudes is proposed.

4.1 Assumptions

I wish to start with a number of assumptions that will be taken for granted in what follows, although no developed argument will be given for their sake in this paper.[13]

Assumption 1. Predicates of personal taste are context-sensitive expressions that express different semantic contents relative to different contexts of utterance.

This idea is endorsed by a number of hybrid taste-expressivists—as well as many contextualists that are not (hybrid) taste-expressivists.[14] Using Kaplanian terminological orthodoxy, taste predicates have both character and content;

13 Assumptions 1 and 2 are defended in Zouhar (2018).

14 See, for example, Buekens (2011), Gutzmann (2016), López de Sa (2008), and Sundell (2011).

character is a non-constant function from contexts of utterance to contents, and content is a function from circumstances of evaluation to extensions. It bears noting that characters of taste predicates are *non*-constant functions; due to this fact, taste predicates closely resemble ordinary indexicals. Content can be identified with a property that is understood as an intension mapping possible worlds to sets of individuals that instantiate the property in question. Based on this, taste predicates, when uttered by different speakers, may express different properties. Clearly, independently of any context of utterance taste predicates express no properties—they only have characters.

Assumption 2. For all values of *x* (where *x* ranges over agents) it holds that, when uttered relative to *x*'s perspective, the predicate 'delicious' expresses, loosely speaking, the property of *interacting with x's gustatory capacities in a way that produces agreeable gustatory experiences of an intrinsically desirable sort.*[15]

This analysis of taste properties is very close to Egan's (see, in particular, Egan 2010); in fact, it is almost identical to Egan's, but with one important difference: Egan proposed to take taste properties as dispositional, whereas the analysis given above is not cast in dispositional terms.[16] I have argued for this amendment elsewhere. What is crucial for the above suggestion is that it nicely coheres with Assumption 1 because it permits to treat taste properties as context-dependent. If Alice utters 'delicious' (and adopts an autocentric perspective), she expresses the property of *interacting with Alice's gustatory capacities in a way that produces agreeable gustatory experiences of an intrinsically desirable sort*; if Bert utters the predicate (while adopting an autocentric perspective), he expresses the property of *interacting with Bert's gustatory capacities in a way that produces agreeable gustatory experiences of an intrinsically desirable sort*. Thus, depending on whose perspective is in play, 'delicious' expresses different properties relative to different contexts of utterance. Apart from this, the above analysis can be recommended for other reasons as well. It suggests that deliciousness is determined by the interactions with an agent's gustatory capacities. It also suggests that the results of these interactions are agreeable experiences—of course, things that are not agreeable may not be delicious. Finally, these gustatory experiences are described as intrinsically desirable. Again, if the experi-

15 In what follows, *x* is supposed to range over agents (I will not mention this fact anymore). If *x* is left unbound by an explicit quantifier, it should be understood as representing an unspecified agent that is *x*'s value. I distinguish between being an agent and being a speaker—the set of speakers is just a subset of the set of agents.

16 The dispositional treatment of taste properties is modelled on Lewis's theory of values; see Lewis (1989). For some other accounts of taste properties along the same lines see Clapp (2015) and López de Sa (2017).

ence is undesired, an agent may not describe the object that caused it as delicious. All these features seem to be rather intuitive. In what follows, I will abbreviate the property of *interacting with x's gustatory capacities in a way that produces agreeable gustatory experiences of an intrinsically desirable sort* as *being delicious*$_x$.

Assumption 3. For every property there is a set of properties that are its requisites.

Generally speaking, property R is a *requisite* of property P if and only if for all possible worlds w and objects y it holds that if y instantiates P at w it also instantiates R at w.[17] It means that requisites of a certain property are all those properties that are necessarily co-instantiated with the given property. For example the property of *being a philosopher* has as its requisites the properties of *being human*, *being a sentient being*, *being rational*, etc. Let us assume that proposition p contains property P as its constituent and that P has as its requisites properties R_1, ..., R_n. It is obvious that if p is true then all propositions r_1, ..., r_n also are true, where r_i results from p by replacing P with R_i ($1 \leq i \leq n$). Moreover, if someone sincerely utters a sentence that expresses p as its semantic content, one suggests that p is true and thereby implies that r_1, ..., r_n also are true. The requisite relation between properties is not empirical—it is independent of any contingent matters of fact that R_1, ..., R_n are requisites of P. If an individual is a philosopher, the individual is e. g. human not just contingently, but necessarily, in the sense that if the individual were not human she could not have been a philosopher either. At the epistemic level, if someone is fully competent with the property of *being a philosopher*, one knows (at least tacitly) a number of other propositions, including the propositions *that only human beings can be philosophers*, *that all philosophers are rational*, and *that there cannot be philosophers who are not sentient beings*. Of course, I admit that full competence can be rare and that partial competence prevails, meaning that someone may know that R_i is a requisite of P, but fail to know that R_j is a requisite of P (where $i \neq j$). Nevertheless, this does not affect the fact that both R_i and R_j are requisites of P.

4.2 Requisites of Taste Properties and Implied Propositions

The notion of requisite is essential to the explanation of non-doxastic taste disagreements proposed in this paper. The first step in showing this consists in rec-

17 The notion of requisite invoked here is adapted from Duží *et al.* (2010: 361). The requisite relation between properties is reflexive and transitive; see (*ibid:* 363).

ognizing that taste properties have certain attitudinal properties as their requisites. In particular, for any value of *x*, the property of *being delicious*$_x$ can be said to have the property of *having a taste that is liked by x* as its requisite. The latter property is attitudinal because some object *y* instantiates it only provided *x* bears the attitude of liking toward *y*'s taste. In some cases, the relevant agent of the context of utterance is the speaker; this happens when the speaker adopts an autocentric perspective, and describes a certain object as delicious from her own perspective. In some other cases, the relevant agent is someone else; in particular, if the speaker adopts an exocentric perspective, the relevant agent is an addressee or a third party. Clearly, the role of agent can be played by a whole group of people; some groups are such that speakers are their members, and some are such that speakers are not their members.

The following observations seem to be a rather straightforward consequence of the idea that taste properties have suitable attitudinal properties as their requisites:

1. There are taste properties that have the same attitudinal properties as their requisites and, yet, are different from one another. For example, both *being delicious*$_x$ and *being tasty*$_x$ have *having a taste that is liked by x* as a requisite. They are different properties but closely similar in the sense that they can be simultaneously exemplified by the same object.[18]
2. Incompatible taste properties have, as a rule, incompatible attitudinal properties as their requisites, where two properties are incompatible provided if an object instantiates one of them it cannot instantiate the other. For example, *having a taste that is liked by x* is a requisite of *being delicious*$_x$ and *having a taste that is disliked by x* is a requisite of *being unpalatable*$_x$, where *being delicious*$_x$ and *being unpalatable*$_x$ on one hand and *having a taste that is liked by x* and *having a taste that is disliked by x* on the other are pairs of incompatible properties.
3. Importantly, an object instantiates a taste property only provided there is an agent who bears a certain kind of evaluative attitude toward the object. This is because instantiating a taste property implies instantiating an appropriate attitudinal property, and an object may instantiate the latter only provided a relevant person bears a certain kind of evaluative attitude toward it. For example, an object cannot instantiate *being delicious*$_x$ unless *x* is such that she likes the object's taste.

18 Moreover, the two properties are not independent because, intuitively, if an object instantiates the property of *being delicious*$_x$ it has to instantiate also the property of *being tasty*$_x$ (but not *vice versa*).

4. A closely related point is that the compatibility of taste properties is connected with the cotenability of the related evaluative attitudes toward objects—if an object instantiates compatible taste properties then there is an agent who bears cotenable evaluative attitudes toward the object (provided she bears different attitudes toward it). On the other hand, if it happens that an object is supposed to instantiate incompatible taste properties then there is an agent who bears non-cotenable evaluative attitudes toward the object; such an agent suffers from having an inconsistent viewpoint.[19]

The claim that taste properties have attitudinal properties as their requisites should be vindicated, though it undoubtedly appears to be rather natural. Why should we say that the taste property of *being delicious*$_x$ and the attitudinal property of *having a taste that is liked by x* are relata of the requisite relation? The explanation can be easily derived from (*i*) the analysis of taste properties according to which the property of *being delicious*$_x$ is unpacked as the property of *interacting with x's gustatory capacities in a way that produces agreeable gustatory experiences of an intrinsically desirable sort*, and (*ii*) the claim that an object instantiates the property of *being delicious*$_x$ only provided the agent *x* bears the evaluative attitude of liking toward the object's taste. Based on (*i*), delicious things are supposed to produce agreeable gustatory experiences that are desirable for the experiencer. If an agent has agreeable gustatory experiences of a desirable sort with a certain food she surely likes the taste of that food, i. e. she bears a particular evaluative attitude toward its taste. In other words, producing gustatory experiences that are both agreeable and desirable is incompatible with the agent being such that she does not bear this kind of evaluative attitude toward the taste of the food in question. Basically, this is the moral that is to be derived from (*ii*) in connection with (*i*). Now if a food is such that an agent likes its taste, i. e. bears a certain evaluative attitude toward it, then, trivially, the food instantiates the property of having a taste that is liked by the agent. Hence, whenever an object instantiates the property of *being delicious*$_x$ it likewise instantiates the property of *having a taste that is liked by x*.

19 Two remarks are to be explicitly made on this last point. First, recall that if an object instantiates compatible taste properties, in some cases an agent may take just one evaluative attitude toward it (see point 1). Thus, the first sentence of point 4 is to be restricted to situations in which agents take different evaluative attitudes toward objects that instantiate compatible taste properties. Second, the properties of *being delicious*$_x$ and *being unpalatable*$_z$ (where both x and z range over agents, and $x \neq z$) are not incompatible. The second sentence is thus to be understood as dealing with situations in which just one agent is involved.

As I have already said, if p is a proposition that contains property P as its constituent and R_1, ..., R_n are P's requisites, then the following holds: if p is true then all propositions r_1, ..., r_n (where for any i such that $1 \leq i \leq n$ it holds that r_i results from p by replacing P with R_i) also are true. Based on this, if the speaker asserts a sentence that expresses p relative to her context of utterance, she expresses a proposition that implies all propositions r_1, ..., r_n. For the sake of simplicity, I will say that by expressing a certain proposition (such as p) the speaker implies some other proposition (such as r_i); however, this is not to be understood as suggesting that the propositions implied are Gricean implicatures or some other pragmatically obtained propositions. Instead, the relation of implication between p and r_1, ..., r_n is based on there being a necessary connection between instantiating P and instantiating R_1, ..., R_n.

If P is a taste property and someone sincerely utters a sentence that expresses p (that contains P) as its semantic content relative to a given context of utterance, one suggests that p is true and implies thereby that r_1, ..., r_n also are true. For example, it seems intuitive that the property *being edible*$_x$ is a requisite of *being delicious*$_x$ for any value of x. If Alice sincerely utters a sentence that expresses, relative to her context of utterance, the proposition *that Choco is delicious*$_{Alice}$, Alice thereby implies the proposition *that Choco is edible*$_{Alice}$. The same holds for propositions that contain other requisites of the property of *being delicious*$_x$ as their constituents. Most notably, if Alice utters a sentence that expresses the proposition *that Choco is delicious*$_{Alice}$, she implies the proposition *that Choco has a taste that is liked by Alice*. In general, the propositions that contain taste properties as their constituents imply (in the above sense) the propositions that contain suitable attitudinal properties in place of the taste properties. The latter will be called *attitudinal propositions* from now on. By saying that an attitudinal proposition is relevant with respect to a taste proposition I mean to say that the former is related to the latter on the basis of there being the requisite relation between the attitudinal property contained in the former and the taste property contained in the latter.[20]

20 Taste disagreements are explained in terms of implied attitudinal propositions. Now, given the fact (alluded to in 4.1) that one may not be aware of the propositions r_1, ..., r_n that are implied by p in the sense described in the main text, it might happen that two speakers express taste propositions that imply incompatible attitudinal propositions without actually being aware of the implied propositions and, consequently, without being aware of disagreeing with each other. It thus seems that the requisite-based account admits of situations in which people may disagree about taste without actually being aware of their disagreement. This seems to be rather counterintuitive. (I am grateful to an anonymous reviewer for raising this challenge.) I admit that the requisite-based account makes theoretical room for such situations. However, in practice, it cannot occur

4.3 Non-Doxastic Disagreement about Taste

Taste propositions imply attitudinal propositions by default. Since attitudinal propositions that are implied by taste propositions are true only provided the agents that are relevant relative to the taste propositions bear appropriate evaluative attitudes toward the things mentioned therein, by asserting her taste proposition a speaker implies bearing a particular evaluative attitude toward the thing in question, and, by so doing, manifests an evaluative attitude toward the thing.

This is crucial for the explanation of non-doxastic taste disagreement because instances of non-doxastic disagreements about taste are supposed to arise when the agents adopt non-cotenable evaluative attitudes toward the same thing. Based on this, if a speaker asserts a taste proposition she thereby implies bearing a particular evaluative attitude toward the food in question. This is not to be understood as meaning that the speaker performs two separate activities, namely expressing a proposition and manifesting an attitude; rather, she performs the former kind of activity, and the latter kind of activity is thereby obtained as a sort of free lunch, so to speak.

This fact points to the main difference between the hybrid taste-expressivist approach and the approach adopted in this paper. Hybrid taste-expressivism has it that speakers both express certain taste propositions and manifest evaluative attitudes toward something, i. e. that they perform two kinds of activities that can be separated from each other. This is witnessed by the hybrid taste-expressivists' idea that, in some situations, speakers may perform one act, namely expressing propositions, without performing the other act, namely manifesting evaluative attitudes. On the other hand, the approach promoted in this paper has it that speakers perform just one kind of activity, namely expressing taste propositions; the effects, which the hybrid taste expressivists are supposed to achieve thanks

provided people understand what being delicious (and other taste properties) really amounts to. It may hardly happen that when one sincerely depicts something as delicious one fails to recognize that one likes the taste of the food in question—in fact, liking the food's taste is normally the primary reason for saying about the food that it is delicious. Similarly, when one witnesses someone else's sincere utterance in which something is depicted as delicious one would hardly fail to recognize that the other person likes the taste of the food in question. If such a situation would occur we might suspect that one failed to grasp what being delicious really amounts to. Putting such situations aside as irrelevant, it is actually the case that when *x* sincerely ascribes to something the property of *being delicious*$_x$, *x* cannot but be aware of self-ascribing the property of liking the taste of the thing in question. Based on this, although the requisite-based account makes room for the theoretical option contained in the objection, such a situation would never materialize because of the above requirement that one has to fulfill in order to really know what being delicious (and other taste properties) really amounts to.

to the other activity, are in fact automatically obtained by expressing the taste proposition in question because the taste proposition implies a relevant attitudinal proposition by default.

Based on the intimate connection between taste propositions and attitudinal propositions, if the speaker's utterance expresses a taste proposition her audience may automatically take it for granted that, by ascribing a taste property to something, the speaker manifests an evaluative attitude toward the thing in question; in their response, by ascribing a suitable taste property to the same thing, the audience manifests another evaluative attitude toward the thing, reacting thereby to the speaker's manifestation. Given this point, if the speaker and her audience express taste propositions in which an object is ascribed certain taste properties, the audience can be said to disagree with the speaker if the former manifests an evaluative attitude that is non-cotenable with the attitude manifested by the latter. This approach can be applied to all kinds of situations, including those in which speakers and their audiences adopt exocentric perspectives, because it is impossible to express a taste proposition without implying a relevant attitudinal proposition.

A general explanation that covers both the situations that involve autocentric perspectives and the situations that involve exocentric perspectives can be formulated along the following lines:

Assume that (*i*) S_1 and S_2 are speakers, (*ii*) A_1 and A_2 are agents, (*iii*) S_1 utters a taste sentence that expresses a taste proposition P_1 that is true about A_1, and (*iv*) S_2 utters a taste sentence that expresses a taste proposition P_2 that is true about A_2. By uttering her taste sentence that expresses P_2 in response to S_1's utterance of a taste sentence that expresses P_1, S_2 presents A_2 as disagreeing with A_1 provided (*i*) P_2 implies an attitudinal proposition according to which A_2 takes a particular evaluative attitude T_2 toward a certain thing, (*ii*) P_1 implies an attitudinal proposition according to which A_1 takes a particular evaluative attitude T_1 toward the same thing, and (*iii*) T_2 is non-cotenable with T_1.

Clearly, the speakers adopt autocentric perspectives provided $S_1 = A_1$ and $S_2 = A_2$, and they take up exocentric perspectives provided $S_1 \neq A_1$ and $S_2 \neq A_2$.

Let us illustrate the prospects of this explanation in terms of the exchange captured in (1). Assuming that 'this chocolate cake' designates Choco relative to both Alice's and Bert's contexts of utterance and that the semantic content of complex demonstratives is exhausted by designated objects,[21] Alice expresses

21 I do not mean to suggest that this approach to the semantic content of complex demonstratives is correct. I adopt this account just because it is rather simple. Nothing important hinges on whether the semantic theory of complex demonstratives that is invoked in the context of the present discussion is correct or not.

the proposition *that Choco is delicious*$_{Alice}$ by asserting 'This chocolate cake is delicious'. Let us assume that her assertion is sincere and that she literally means what she says. Alice knows that Choco tastes good to her; otherwise she would hardly want to express the proposition *that Choco is delicious*$_{Alice}$ in a serious manner. Her audience Bert also knows that people sincerely and literally say about something that it is delicious only provided they like the taste of the food in question. Thus, encountering Alice's sincere and literal utterance of 'This chocolate cake is delicious', he recognizes that she likes the taste of Choco. At the same time, however, he is aware of his disliking the taste of Choco. He knows that he would succeed in manifesting this attitude by uttering a sentence that is a linguistic negation of Alice's sentence. With this aim in mind, he utters 'This chocolate cake is not delicious', expressing thus the proposition *that Choco is not delicious*$_{Bert}$. In so doing, he manifests an attitude that is non-cotenable with the attitude manifested by Alice. Despite the fact that the propositions *that Choco is delicious*$_{Alice}$ and *that Choco is not delicious*$_{Bert}$ are compatible, they imply attitudinal propositions that contain non-cotenable evaluative attitudes as their constituents. As a result, Alice and Bert present themselves as taking non-cotenable attitudes toward Choco. That is why he disagrees with her; their disagreement is thus non-doxastic.

This kind of explanation can be easily extended to the problematic cases discussed in Section 3. First, consider the exchange captured in (2) that involves an exocentric perspective. If Alice utters 'Choco is delicious' while having in mind Cynthia's perspective, and thereby expresses the taste proposition *that Choco is delicious*$_{Cynthia}$, the attitudinal proposition *that Choco has a taste that is liked by Cynthia* is automatically implied. Alice thus succeeds in suggesting which particular evaluative attitude Cynthia takes toward Choco. Alice succeeds in this because the two propositions are connected on the basis of the requisite relation between the properties contained in them. Bert responds by uttering 'Choco is not delicious' while having in mind Dave's perspective. His utterance expresses the proposition *that Choco is not delicious*$_{Dave}$ and, in so doing, implies the proposition *that Choco has a taste that is disliked by Dave*. In this case, it is thus Dave's evaluative attitude toward the cake that is made obvious. Since the two evaluative attitudes are non-cotenable and both Cynthia and Dave are presented as taking these attitudes toward the same food, Dave's reaction to the taste of Choco is interpreted as disagreeing with Cynthia's reaction to it. Of course, there is no disagreement between Alice and Bert over the taste of Choco; their respective utterances are to be understood as records of the disagreement that arises between Cynthia's and Dave's evaluative attitudes toward Choco. As a result, Alice and Bert present Cynthia's and Dave's respective responses to Choco as an instance of non-doxastic taste disagreement.

The second problematic kind of case arises with respect to utterances of embedded taste sentences. The dialogue captured in (3) is an instance of this kind. Given the requisite-based account, what is important is that the utterances of embedded sentences express taste propositions and, despite being embedded, imply relevant attitudinal propositions. Alice's utterance of 'It is delicious', in which 'it' occurs as anaphorically dependent on 'Choco', expresses the proposition *that Choco is delicious*$_{Alice}$; Alice suggests that the truth of this proposition is conditionally dependent on the truth of the proposition *that Cynthia finds Choco delicious.*[22] Based on this, she suggests that she takes the evaluative attitude of liking toward Choco's taste conditionally relative to the latter proposition's being true. Bert responds by suggesting that if the condition according to which Cynthia finds Choco delicious is fulfilled, the proposition *that Choco is not delicious*$_{Bert}$ would be true. In so doing, he makes plain that he would adopt the evaluative attitude of dislike toward Choco's taste provided Cynthia's attitude toward it is appreciative. Thus, this case contains a kind of disagreement between Alice and Bert that consists in the following: if a certain condition (that concerns Cynthia's attitude toward the taste of Choco) obtains, there arises a situation in which Alice and Bert adopt non-cotenable evaluative attitudes toward the taste of Choco. The mutual compatibility of the proposition *that Choco is delicious*$_{Alice}$ and the proposition *that Choco is not delicious*$_{Bert}$ notwithstanding, Bert obviously disagrees with Alice. To conclude, the requisite-based account provides a natural explanation of this conditional kind of taste disagreement in non-doxastic terms.

5 Conclusion

The requisite-based account of non-doxastic taste disagreement presented in this paper seems to be more general than some hybrid taste-expressivist accounts that can be found in the literature. The former account is able to deal successfully with certain kinds of situations that the latter accounts cannot explain in a satisfactory manner.

The main difference between the two kinds of accounts consists in that while the requisite-based approach assumes that there is a stable connection between taste propositions and attitudinal propositions that holds independently of any context of utterance (as well as irrespective of the speakers' communicative intentions and objectives), the hybrid taste-expressivist accounts take it that expressing

22 There is a natural reading according to which Alice's utterance of 'It is delicious' expresses the proposition *that Choco is delicious*$_{Cynthia}$. This reading is irrelevant in the present situation.

taste propositions can be detached from manifesting evaluative attitudes toward something. There are at least two kinds of situations in which speakers express taste propositions without manifesting evaluative attitudes toward something—the cases in which they take up exocentric perspectives and the cases in which they utter taste sentences as embedded within larger linguistic constructions of certain kinds. Hybrid taste expressivism cannot describe these cases as instances of non-doxastic taste disagreement because this kind of disagreement is claimed to arise only provided the speakers manifest their evaluative attitudes toward a particular thing by using suitable locutions in an expressive way.

The account proposed in this paper does not make the instances of non-doxastic taste disagreement dependent on expressive uses of taste sentences. That is why it also can handle the problematic situations in which no expressive uses occur. Despite the absence of expressively used taste sentences, the speakers succeed in manifesting evaluative attitudes that are the basis for expressing their disagreement as non-doxastic. Consequently, the main contribution of the present account consists in that it divorces the idea that taste disagreements are best explained as non-doxastic from the idea that speakers disagree with their interlocutors in the non-doxastic sense provided their taste sentences are uttered in the expressive way.[23]

References

Baker, Carl. 2012. Indexical Contextualism and the Challenges from Disagreement. *Philosophical Studies* 157(1). 107–123.

Barker, Chris. 2013. Negotiating Taste. *Inquiry* 56(2–3). 240–257.

Barker, Stephen. 2010. "Cognitive Expressivism, Faultless Disagreement, and Absolute but Non-Objective Truth". *Proceedings of the Aristotelian Society* 110(2). 183–199.

Boisvert, Daniel. 2008. "Expressive-Assertivism". *Pacific Philosophical Quarterly* 89(2). 169–203.

Boisvert, Daniel. 2014. "Expressivism, Nondeclaratives, and Success-Conditional Semantics". In Guy Fletcher and Michael Ridge (eds.), *Having It Both Ways: Hybrid Theories and Modern Metaethics*. Oxford: Oxford University Press. 22–50.

Buekens, Filip. 2011. "Faultless Disagreement, Assertions and the Affective-Expressive Dimension of Judgments of Taste". *Philosophia* 39(4). 637–655.

Cappelen, Herman and Hawthorne John. 2009. *Relativism and Monadic Truth*. Oxford: Oxford University Press.

Clapp, Lenny. 2015. "A Non-Alethic Approach to Faultless Disagreement". *Dialectica* 69(4). 517–550.

23 I am indebted to an anonymous reviewer for valuable suggestions that helped me to improve the paper in several respects. This paper was supported by VEGA grant No. 1/0197/20.

Copp, David. 2001. "Realist-Expressivism: A Neglected Option for Modal Realism". *Social Philosophy and Policy* 18(2). 1–43.
Copp, David. 2014. "Can a Hybrid Theory Have It Both Ways? Moral Thought, Open Questions, and Moral Motivation". In Guy Fletcher and Michael Ridge (eds.), *Having It Both Ways: Hybrid Theories and Modern Metaethics*. 51–74. Oxford: Oxford University Press.
Duží, Marie, Jespersen Bjørn and Materna Pavel. 2010. *Procedural Semantics for Hyperintensional Logic: Foundations and Applications of Transparent Intensional Logic*. Dordrecht: Springer.
Egan, Andy. 2010. "Disputing about Taste". In Richard Feldman and Ted A. Warfield (eds.) *Disagreement*. 247–286. Oxford: Oxford University Press.
Egan, Andy. 2014. "There's Something Funny about Comedy: A Case Study in Faultless Disagreement". *Erkenntnis* 79(1). 73–100.
Eriksson, John. 2014. "Hybrid Expressivism: How to Think about Meaning". In Guy Fletcher and Michael Ridge (eds.) *Having It Both Ways: Hybrid Theories and Modern Metaethics*. 149–170. Oxford: Oxford University Press.
Glanzberg, Michael. 2007. "Context, Content, and Relativism". *Philosophical Studies* 136(1). 1–29.
Gutzmann, Daniel. 2015. *Use-Conditional Meaning: Studies in Multidimensional Semantics*. Oxford: Oxford University Press.
Gutzmann, Daniel. 2016. "If Expressivism Is Fun, Go for It!". In Cécile Meier and Janneke van Wijnberger-Huitink (eds.) *Subjective Meaning: Alternatives to Relativism*. 21–46. Berlin: de Gruyter.
Hirvonen, Sanna. 2016. "Doing without Judge Dependence". In Cécile Meier and Janneke van Wijnberger-Huitink (eds.) *Subjective Meaning: Alternatives to Relativism*. 47–68. Berlin: de Gruyter.
Huvenes, Torfinn T. 2012. "Varieties of Disagreement and Predicates of Taste". *Australasian Journal of Philosophy* 90(1). 167–181.
Huvenes, Torfinn T. 2014. "Disagreement without Error". *Erkenntnis* 79(1). 143–154.
Iacona, Andrea. 2008. "Faultless or Disagreement". In Manuel García-Carpintero and Max Kölbel (eds.) *Relative Truth*. 287–296. Oxford: Oxford University Press.
Kölbel, Max. 2002. *Truth without Objectivity*. London—New York: Routledge.
Kölbel, Max. 2003. "Faultless Disagreement". *Proceedings of the Aristotelian Society* 104(1): 53–73.
Lasersohn, Peter. 2005. "Context Dependence, Disagreement, and Predicates of Personal Taste". *Linguistics and Philosophy* 28(6). 643–686.
Lewis, David. 1989. "Dispositional Theories of Value". *Proceedings of the Aristotelian Society* 63. 113–138.
López de Sa, Dan. 2008. "Presuppositions of Commonality: An Indexical Relativist Account of Disagreement". In Manuel García-Carpintero and Max Kölbel (eds.) *Relative Truth*. 297–310. Oxford: Oxford University Press.
López de Sa, Dan. 2017. "Making Beautiful Truths". In James O. Young (ed.), *Semantics of Aesthetic Judgements*. 38–60. Oxford: Oxford University Press.
MacFarlane, John. 2014. *Assessment Sensitivity: Relative Truth and its Applications*. Oxford: Oxford University Press.

Marques, Teresa. 2016. “Aesthetic Predicates: A Hybrid Dispositional Account”. *Inquiry* 59(6). 723–751.
Marques, Teresa and García-Carpintero Manuel. 2014. “Disagreement about Taste: Commonality Presuppositions and Coordination”. *Australasian Journal of Philosophy* 92(4). 701–723.
Palmira, Michele. 2015. “The Semantic Significance of Faultless Disagreement”. *Pacific Philosophical Quarterly* 96(3). 349–371.
Plunkett, David and Sundell Tim. 2013. “Disagreement and the Semantics of Normative and Evaluative Terms”. *Philosopher’s Imprint* 13(23).
Recanati, François. 2007. *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.
Richard, Mark. 2008. *When Truth Gives Out*. Oxford: Oxford University Press.
Rosenkranz, Sven. 2008. “Frege, Relativism and Faultless Disagreement”. In Manuel García-Carpintero and Max Kölbel (eds.) *Relative Truth*. 225–238. Oxford: Oxford University Press.
Schaffer, Jonathan. 2011. “Perspective in Taste Predicate and Epistemic Modals”. In Andy Egan and Brian Weatherson (eds.) *Epistemic Modality*. 179–226. Oxford: Oxford University Press.
Silk, Alex. 2016. *Discourse Contextualism: A Framework for Contextualist Semantics and Pragmatics*. Oxford: Oxford University Press.
Smith, Barry C. 2010. “Relativism, Disagreement and Predicates of Personal Taste”. In François Recanati, Isidora Stojanovic & Neftalí Villanueva (eds.), *Context-Dependence, Perspective and Relativity*. 195–223. Berlin—New York: de Gruyter.
Stojanovic, Isidora. 2007. “Talking about Taste: Disagreement, Implicit Arguments, and Relative Truth”. *Linguistics and Philosophy* 30(6). 691–706.
Stojanovic, Isidora. 2011. “When (True) Disagreement Gives Out”. *Croatian Journal of Philosophy* 11(2). 183–195.
Sundell, Tim. 2011. “Disagreements about Taste”. *Philosophical Studies*. 155(2). 267–288.
Wright, Crispin. 1992. *Truth and Objectivity*. Cambridge, (MA): Harvard University Press.
Wyatt, Jeremy. 2018. “Absolutely Tasty: An Examination of Predicates of Personal Taste and Faultless Disagreement”. *Inquiry* 61(3). 252–280.
Zeman, Dan. 2016. “Contextualism and Disagreement about Taste”. In Cécile Meier and Janneke van Wijnberger-Huitink (eds.) *Subjective Meaning: Alternatives to Relativism*. 91–104. Berlin: de Gruyter.
Zeman, Dan. 2017. “Contextualist Answers to the Challenge from Disagreement”. *Phenomenology and Mind* 12. 62–73.
Zouhar, Marián. 2014. “In Search of Faultless Disagreement”. *Prolegomena* 13(2). 335–350.
Zouhar, Marián. 2018. “Conversations about Taste, Contextualism, and Non-Doxastic Attitudes”. *Philosophical Papers* 47(3). 429–460.

Joanna Klimczyk

Why the Basic Problem Is Not a Problem

Abstract: The paper discusses Mark Schroeder's famous objection to the interpretation of agentive ought in terms of propositional ought. The objection in question is called the Basic Problem, and it amounts to the observation that systematic, uniform logical treatment of 'ought' sentences that express deliberative content and those that do not gives rise to a semantic anomaly. The anomaly in question is that we are forced to accept as meaningful sentences that obviously lack meaningfulness. If Schroeder's argument is a good one, then the challenge is truly fatal. However, I argue, this is not so. The charge looks very serious, I argue, once we accept two assumptions. One is the assumption that makes the Basic Problem *a* problem, namely that some normative thoughts are unthinkable. The other is the assumption that gives the Basic Problem the appearance of being devastating and irrefutable, and amounts to the idea that there is a tight conceptual relationship between a triple of concepts 'indexed ought', 'normatively owned ought', and 'obligation'. I shall show that both assumptions do not withstand criticism. If my arguments are correct, the Basic Problem turns out to be toothless.

1 Introduction

The paper is divided into four sections. In section one I explain what the Basic Problem is, and why it has every appearance of being fatal for a proponent of the unifying view—the view saying that any normative ought-content, including the genuinely agentive one, can be expressed in terms of propositional ought. In section two, I focus on undermining Schroeder's initial idea that certain normative thoughts are truly unthinkable as they are either empty or have content that we are unable to make sense of, which finally makes their content cognitively un-

The paper was written under support of the National Science Centre from the funds in grant OPUS 12 on the project 'Substantive semantics for normative language (on the basis of the analysis of 'ought' sentences). Decision no. DEC-2016/23/B/HS1/02921. Thanks to Arkadiusz Chrudzimski for his generous comments on an earlier version of the paper, and to an anonymous referee for their helpful remarks.

Joanna Klimczyk, Institute of Philosophy and Sociology of the Polish Academy of Sciences, Warsaw (jklimczyk@ifispan.edu.pl)

https://doi.org/10.1515/9783110702286-005

available, thence empty *for us*. To that aim I sketch a toy position in the theory of normative concepts that shows that Schroeder is mistaken in thinking that gross departure from the grammar of 'ought' renders 'ought' sentences unintelligible, because heavily deviant 'ought' sentences cannot express any thought. Drawing on the results of my critical examination of the *empty thought thesis*, as I call it, in section three I discuss the tacit assumption that lies in the background of the whole debate over the proper logical and grammatical interpretation of 'ought' sentences with truly agentive content. Here it is argued that the assumption in question is a heuristic principle that posits an incredible conceptual connection between the triple of concepts, which are 'indexed ought', 'normatively owned ought' and 'obligation'. I consider the principle in question under the guise of the general thesis saying that indexing 'ought' to a subject is a reliable guide to that 'ought's' normative meaning. Section four switches the critical perspective from external to internal, and is targeted at vindicating the conditional claim that *if*, contrary to the presented arguments, the Basic Problem remains in force, then Schroeder's own interpretation of genuinely agentive ought in terms of a normative relation between an agent and an action, construed as a sort of property of the former, is also affected by it. That is, I argue to the effect that if the Basic Problem does any damage, it does damage across the board.

2 The Basic Problem at its Basis

In his influential 2011 paper entitled "'Oughts', Agents and Action", Mark Schroeder presents a rigorous and at the same time vigorous criticism of the dominant view in the philosophical approach to normative semantics of 'ought' sentences. The dominant view hangs heavily on the work in linguistic semantics and deontic logic, and boils down to the claim that *any* normative content can be expressed by 'ought' sentences in which 'ought' is interpreted as a propositional operator. And the only question that remains to be answered concerns the issue of how much, if at all, enhancement we do need to introduce into the grammatical interpretation of agentive[1] ought, if it is still to be accommodated within the propositionalist framework (cf. Broome 2013, Wedgwood 2006, 2007, Chrisman 2015). Agentive oughts, i. e. oughts typically expressed by sentences of the form 'S ought to φ' and interpreted as saying what *S herself* is required

1 Throughout the paper I stick to the following terminological regime: I use the expression 'agentive ought' to refer to ought being normally interpreted in terms of personal requirement or simply as a genuine 'ought to do'.

to do, pose a problem for the interpretation in terms of propositional ought, since English grammar is too coarse-grained to make it crystal clear that the agent related to the relevant proposition is to be the sole agent involved in making that proposition true.

According to the dominant view, irrespectively of the sort of normative claim you make (epistemic, moral, practical, etc.), the very message your judgment is purported to carry can be successfully represented by a sentence of the form *O*(p) or by its enhanced version A*O*(p).[2] Following Finlay & Snedegar 2014, let us think about the dominant view under the guise of the *Uniformity Thesis* (*UT* for short), saying the following: 'ought' has unified semantics and logical form (syntax). According to *UT,* 'ought' sentences expressing agentive content are semantically and logically equivalent to 'ought' sentences expressing non-agentive content. The essence of *UT* boils down to the idea that 'ought' always operates on proposition.

UT is an orthodox unifying stance in that it states explicitly that agential 'ought' sentences with evident agentive content such as 'Peter ought to brush his teeth at least twice a day' are semantically and logically equivalent to non-agential sentences like 'It ought to be that Peter brushes his teeth at least twice a day'[3]. So, according to *UT*, no matter what normative thought, evaluative or practical, you want to express by means of an 'ought' sentence, the content of that thought is appropriately linguistically interpreted in terms of a proposition that ought to obtain.

However, there is also an unorthodox approach to the semantics of 'ought' within the propositionalist camp advocated by Broome 2013 and Wedgwood 2006, 2007 to give just two examples, in which *some* 'oughts', precisely those that are central to everyday life—'truly practical oughts'[4] as I name them, cannot be fully happily represented within the framework of *UT*. This is because the essence of these oughts consists in the fact of their issuing requirements for those to whom they 'belong'. Following the tradition of deontic logic Broome calls them "owned oughts" (2013:13). 'Owned oughts' are standardly interpreted in terms of indexed propositional ought of the following general form *AO*(*p*)

2 Cf. Broome 2013.

3 The unorthodox version of *UT* advanced in Broome 2013 proposes fine-graining on the propositional content of the sentence by inflicting some violence on English grammar. In Broome's proposal, an 'ought' sentence with clearly agentive content, telling the agent what she ought to do, is to be properly grammatically interpreted to the following, general effect '*A* ought that *A* φ-s'.

4 I introduce and motivate this name in Substantive semantics for discourse-relevant 'ought to do', book manuscript. See also Klimczyk 2017.

where '*A*' stands for 'agent', '*O*' is deontic operator and '*p*' the proposition that *A* stands in normative relation to.

Take our toy example which is the sentence 'Peter ought to brush his teeth regularly'. If this sentence's natural interpretation is the one in which Peter is required to be the agent of brushing his teeth, therefore the ought expressed by the sentence is 'owned' in Broome's terminology, then the sentence's proper interpretation is this: 'Peter ought that Peter brushes his teeth'[5].

Now, Schroeder's challenge seems to be primarily targeted at the refined propositionalist interpretation (*RI* for short) of agential 'ought' sentences, in which 'ought' is taken to relate an agent to a proposition. He invites us to examine what follows had *RI* been generalized, and the formula *AO*(p) served as a matrix of interpretation of *any* 'ought' sentence with alleged agentive content. At first blush, such a generalization would be much welcome, because indexed 'ought' better serves the role of an indicator of owned ought than the unindexed 'ought' does. However, this obvious advantage aside, *RI* generates a true worry which is that it puts no constraints on *what* action-proposition *A* can stand in relation to. Crucially, it remains silent about the *sort* of action-proposition that fits the extension of 'proposition to which the agent is to be sensibly related'. In any case, it is obvious that the correct *RI* of an agentive ought cannot be nonsensical. And if *RI* of agentive ought is to make sense, then there must be a tacit constraint inbuilt into *RI* regarding the sort of action-proposition that an agent *can* stand in normative relation to, namely that the action-proposition describes an activity of the subject of 'ought'. Call the developed version of *RI RI+*.

Though *RI+* nicely delimits the scope of agential 'ought' sentences that are to be bearers of agentive content, it is unsatisfactory when assessed from a theoretical perspective. At least this is how I interpret Schroeder's worry. The problem with *RI+* is that it is *ad hoc*. If truly practical ought ('deliberative' in Schroeder's terminology) relates an agent to an action-proposition, why not assume that it relates the agent to an *arbitrary* action-proposition? Note, that if *RI* is to be taken as a serious proposal of how to logically interpret genuinely practical content, it should properly generalize. By saying that it should properly generalize, I mean that it should tell us how to decipher the proper logical form of 'ought' sentences with agentive content, and that it should serve as a sort of a guide about how to build *novel* and semantically meaningful 'ought' claims. Put differently, a good *theory* of the interpretation of 'ought' sentences with agentive content should be both *backward-looking* (should explain how to logically interpret

5 I discuss Broome's grammatical interpretation of practical ought at length in *Substantive semantics*.

encountered normative sentences of the relevant sort), and *forward-looking (*should offer a clue about what normative content can be expressed by 'ought' sentences of the relevant logical structure).

If the challenge presented under the objection from *proper generalization* is on the right track, then both *RI* and its enhanced counterpart *RI+* are to be rejected because *RI* is too powerful and *RI+ ad hoc*. *RI* is too comprehensive since it accommodates all the propositions that it should accommodate—that is, those that *RI+* is concerned with—but it also accommodates those that it *should not* accommodate, precisely those that when related to agents make up nonsensical content. On the other hand, giving up on *RI* in favor of *RI+* is only a seemingly advantageous move, since it turns out that the propositionalist interpretation of 'ought' applies only to some sort of agential 'ought' sentences and not all. Consequently, alluding to *RI+* looks like a desperate way of saving something unrescuable, i.e. the idea that a truly practical ought is to be satisfactorily accountable in terms of the propositional ought, though *in fact* it is not. This is the claim that in my understanding Schroeder's objection is purported to vindicate. Let me now examine his objection in more detail and explain why I think the objection is misfired.

2.1 The Basic Problem

Here is how Schroeder motivates the Basic Problem: "But if OUGHT is just a relation that you can stand in to some proposition—for example, the proposition that you exercise daily—then it is a relation of which it makes sense to ask whether you stand in it to arbitrary other propositions—for example, to the proposition that *I* exercise daily. But I don't think that it makes sense to ask whether you stand in the OUGHT relation to the proposition that I exercise daily. It's not just that I think it is *false* that you ought for me to exercise daily; *I just don't think this question makes any sense; I think it involves a category mistake. So if it is false, it is false because it doesn't make sense*. The view that the OUGHT relation relates agents to propositions is too powerful, because it predicts that some things should make sense which don't—it licenses a category mistake. That is why it is wrong. That is the Basic Problem" (emphasis mine).

As I understand the above passage, the problem that Schroeder draws our attention to is *not* that it is absurd to think, say, that Queen Elizabeth *stands in* some *normative* relation to the proposition *that your uncle will kiss his wife*. The problem in question is more basic than that, and amounts to the mere impossibility of having that very *thought!* If Schroeder is right, you cannot have *that* deviant thought, the thought with the content that Queen Elisabeth ought that

your uncle will kiss his wife. And the reason why, according to him, you cannot entertain the aforementioned thought is *not* that linguistic translation of that thought is syntactically bizarre in the first place, but because it is impossible for one agent to stand in the normative relation to a proposition expressing another agent's activity. Crudely, if I cannot ought for *you* that you φ, nor Queen Elisabeth is in capacity to ought for your uncle that he will kiss his wife, the deliberative issue over whether I or Queen Elisabeth stand in the normative relation to your φ-ing, or your uncle's kissing his wife, does not even arise since normally we do not consider practically unfeasible scenarios.

Obviously, if you cannot think what either of the two considered sentences says, namely that I ought that you exercise daily, or that Queen Elisabeth ought that your uncle will kiss his wife, then obviously you cannot express the content of either of these thoughts linguistically. Empty thought has no representation—linguistic or otherwise. Call this problem *the empty content problem* (*ECP*). *ECP* is the unwelcome consequence of the unifying view that extends the propositional interpretation of non-agential 'ought' to *any* use of 'ought', including the deliberative one. In fact, I contend, it is the *ECP* and not the Basic Problem (henceforth *BP* for short) that poses a true challenge for the propositional interpretation of agentive ought because it is *ECP* and not *BP* that says something which is genuinely hard to accept, namely that our normative thought happens to be empty!

If *ECP* was an unavoidable consequence of *UT*, the Basic Problem would indeed be fatal. Luckily, things do not look so bad. I believe that Schroder is mistaken about a couple of issues which taken together make for his erroneous conviction about the semantic meaninglessness of syntactically deviant agential 'ought' sentences like those considered above. *Firstly*, he is wrong that you cannot think the thought with the cited content. If he is wrong that your thought cannot be *that Queen Elisabeth ought that your uncle will kiss his wife*, then obviously he is wrong that the thought has empty content. So, *secondly*, he is mistaken in thinking that certain odd ought thoughts are unthinkable. *Thirdly*, he is mistaken in his view that grammaticality (the rules of a reputable grammar) delimits the thinkable. But if the rules governing producing meaningful sentences are not to be particularly trustable when delineating thinkable and linguistically expressible content and its boundaries in case of normative 'ought' sentences (which is the claim to be argued for in the paper), there is no good reason to suppose that grammatical correctness (according to a reputable theory) itself is a reliable indicator of semantic meaningfulness.

Now, if ascribing a particular syntactical structure to normative 'ought' sentences does not guarantee ascribing them surface meaningfulness,[6] the general heuristic principle for normative 'ought' that builds upon fostering a tight conceptual relation between certain logical facts—*indexing* 'ought' *to an agent* and its proper interpretation in terms of normative ownership—is untenable. Consequently, and this amounts to my fourth and final critical remark, the *Indexation Implies Normative Ownership Thesis* (*IINOT*), which is an interpretative device widely applied to normative 'ought' sentences with agentive content, is incapable of telling us the normative status of the person who occupies the agent-argument-place. Put differently, my fourth complaint set out under the guise of *IINOT is* the following: even such evidently agentive 'ought' sentences like 'Peter ought to brush his teeth' or 'Ann ought to finish the report' do not warrant the reading in which Peter, or Ann respectively, is the normative and *relevant* bearer of responsibility in case he or she fails to satisfy the ought in question. But if *IINOT* falls short in the role of a reliable heuristic device used to interpret the *normative* meaning of sentences of the above-mentioned sort, *then BP* loses much of its original force. Why care about whether 'ought' can or cannot relate an agent to an arbitrary action-proposition, if from such relating no strong normative consequences follow about the responsibility of the agent in question? Crudely, my point is that, *if* indexing 'ought' to an agent itself is devoid of substantive normative sense, which is the view underlying *IINOT*, then why think that sentences like Schroeder's toy example of bizzare 'ought' sentence 'I ought that you exercise daily' are incomprehensible? Let me now do my best to convince the reader that my scepticism is not unmotivated.

3 Odd Ought Thoughts Are Not So Odd

My first critical remark concerns the stipulated impossibility of entertaining the thought whose transcription in the mental language (which for the purpose of simplicity let's assume is English) is *that Queen Elisabeth ought that your uncle will kiss his wife*. Now, I hold that in order to make the deviant ought-proposition intelligible, what we need is to reject the view that the concept of normative ought has a clear semantic content best interpreted in terms of norms or

6 By 'surface meaningfulness' I mean a linguistic phenomenon taking place when a sentence is well-formed according to the rules of a reliable grammar, and therefore has every appearance of being meaningful, but nevertheless is unintelligible due to the fact that we have no idea as to the meanings of the lexical items in it.

rules that are supposed to make the relevant normative ought-proposition true. In its stead I think we need to adopt a view based on which the semantic content of the concept of normative ought is given by some substantive view on the nature of normativity associated with the concept in question. Call this tentative view *doxastic descriptivism* (or DD for short).[7] Doxastic descriptivism, very roughly, is a novel position in the theory of normative concepts, saying that the content of a normative concept, like the concept of normative ought, is essentially informed by some substantive views on the nature of normativity associated with that particular concept, when it is used as a workaday normative concept. The central idea behind DD is that the concept of normative ought has no single, objective and eternally fixed semantic content, but rather its content is the upshot of one's subscribing to a particular theory, saying what the nature of normative ought is all about. On that doxastically-informed account of the semantic content of normative ought, we are free to say that whether the deviant ought-thought of the considered sort is intelligible or not hangs upon the details of one's substantive views about the intension of the concept in question.

In light of the above stipulations, it turns out that in order to find out whether the proposition *that I ought that you exercise daily* makes any sense, first we have to figure out *what* substantive metanormative content we inscribe into the concept of normative ought. Let me explain. If you assume, as Schroeder does, that the concept of normative ought expressed by the sentences of the form 'S ought to φ' is essentially connected with one's *full* capacity to do the required action (in my terminology this is the substantive content associated with the concept of normative ought, expressed by agentive 'ought' sentences), then Schroeder's toy sentence is unintelligible. But why would we think that this is the only option? Why not suppose that the concept of normative ought is gradable, meaning that one's agency need not be exercised to the full to make one be bound by some ought? Imagine that you find out that your best friend is addicted to some anti-depressant. Is it really bizarre to suppose that, upon discovering that fact, you should say to yourself 'I ought that she will stop taking those pills'? I think it is not. The deviant syntax in which the statement is couched can be interpreted as a mark of the agent's current state of mind: resolution to help her friend *in the situation*, in which he has not yet figured out what *exactly he* ought to do.

To conclude, the general moral to be driven from my brief elucidations on the possibility of having thoughts of the general content being 'S1 ought that S2 φ-s', where 'S1' denotes a distinct person than the one denoted by 'S2', is this: if you generally hold a view that semantic content of abstract philosophical

7 I develop DD in Klimczyk (in progress).

notions like the notion of normative ought is essentially underdetermined and metanormative conception-driven, then nothing stands in the way of your claiming that full capacity to make the ought proposition obtain is not part of *your* concept of normative ought as expressed by sentences of the form 'S ought to φ'. Moreover, if my observation that the intelligibility or not of deviant ought-sentences depends on the intelligibility of a particular theory of normative concepts (in my suggestion, it depends on a particular version of doxastic descriptivism underlying one's understanding of the concept of normative ought) is plausible, then it is not the case that what Schroeder thinks of in terms of deviant ought sentences exist at all.

If these remarks are sensible, they motivate the following tentative conclusion: *if* ECP is to enjoy the reputation of a genuine problem for a proponent of *UT*, then Schroder must provide us with an argument blocking my suggestion about substantive and conception-based semantic content of the concept of normative ought. If I am right, a decisive argument against or for the intelligibility of sentences of the form 'S1 ought that S2 φ-s' is to come from a theory of normative concepts, precisely from establishments in the metaphysics of the concept of normative ought—not, as Schroeder thinks, from the rule book of a reputable grammar.

4 Indexing 'Ought' to an Agent and its Ethical (Non)Significance

In the previous section I stipulated that the challenge that the Basic Problem makes us face is not primarily about the semantic meaningfulness of the alleged deviant 'ought' sentences, but at the bottom points to a fundamental problem in the theory of normative thought. Transposed one level up to the theory of content, the Basic Problem presses us to come to terms with a really disturbing idea that we can have an empty or meaningless normative thought. I called that unwelcome consequence of the Basic Problem the problem of empty content (*ECP*). I have also shown that *ECP* looks serious only *if* we combine the propositionalist nature of normative ought-content with a particular substantive account of agentive normativity on which responsibility is not gradable. If sentences of the general form 'S ought to φ' are to be uniformly read as saying that *S*, being the agent of φ-ing, is in 'the hundred percent' the bearer of responsibility in case she fails to φ, Schroeder's deviant 'ought' sentences *are* deviant—not because they are linguistic translations of thoughts we cannot have, *but* because the conception of normativity that Schroeder associates with the concept of nor-

mative ought, expressed by sentences of a particular grammar, make them inadmissible.

In this section I will launch an attack on the widely accepted interpretative device applied to normative 'ought' sentences that in its most general form is represented under the guise of the *Indexation Implies Normative Interpretation Thesis* (*IINIT*)[8]. I shall argue to the effect that, if *IINIT* is an unreliable principle of interpretation of problematic normative 'ought' sentences, the Basic Problem is toothless. Put differently, I argue that the Basic Problem would be a true problem, if the semantic role accrued to the agent-argument place had been reliably indicative of the normative status of the agent standing in the position of the grammatical subject of 'ought' sentences with *obvious agentive* meaning. Qualifying which agential 'ought' sentences are the litmus test of the considered interpretative device is important because *IINIT* admits of two interpretations: a weaker and a stronger one. At the end of the day both prove false, but they differ in the degree of their falsity, so to speak. The weaker one is a literal interpretation of *IINIT*, in which ascribing 'ought' to the grammatical subject of the sentence is a reliable source of the sentence's normative meaning. The outright falsity of this interpretation consists in the fact that from the grammar of an 'ought' sentence with evident normative content one cannot straightforwardly tell *what normative* proposition it expresses. If normativity is not a unity, and the idea of doxastic descriptivism in the theory of normative concepts is tenable, as I have shown in section two, then it is far from clear exactly what normative proposition the relevant normative 'ought' sentence expresses. And the stronger version of *IINIT* is false, not because it derives its normative conclusion on the basis of the grammatical fact of assigning 'ought' to a subject, but because the *proper* interpretation of an agential 'ought' sentence in terms of agentive ought is unable to settle that the agent to whom ought belongs is also the very agent who is the *normatively relevant* owner of that ought. In a nutshell, the former principle of interpretation of agentive 'ought' sentences is implausible because it inscribes into syntax something that is not there—i.e. substantive semantic content. The latter is implausible because it assumes that the concept of ownership is an outright normative notion of particular content, and then, drawing upon this false assumption, it postulates a tight conceptual relationship between the concept of owned ought (linguistically fleshed out in terms of 'indexed ought') and the concept of ought of obligation.

8 This is inspired by Broome's unnamed principle (p. 14) introduced in his prominent book Rationality through reasoning.

In what follows, I present two arguments against both versions of *IINIT*.[9] The first argument is targeted at the idea that indexation of 'ought' to the agent is anyhow relevant to know that the subject of 'ought' is the normatively relevant owner of that ought. I call the criticized thesis *Indexation Implies Normative Ownership Thesis* (*IINOT*). The second argument, on the other hand, is purported to undermine the seemingly plausible assumption that figuring out the owner of ought is revealing as to who is the normatively relevant owner of that ought. I call the thesis attacked by the second argument *Ownership Implies Obligation Thesis* (*OIOT*). It is important to stress that both theses are inspired by a well-homed idea in deontic logic, i.e. that the notion of ownership is of true help in the analysis of the exact meaning of 'ought' sentences with agentive content.

Here is what the *Indexation Implies Normative Ownership Thesis* (*IINOT* for short). says:

> *IINOT:* indexing[10] 'ought' to a subject of 'ought' in 'ought' sentences of the relevant sort[11] implies that the subject of 'ought' is the normative owner of that ought.

Recall our toy sentence from section one, 'Peter ought to brush his teeth regularly'. Assume next that the context in which that sentence appears makes it clear that the sentence has a clear agentive sense. Now, *IINOT* stipulates that, because the agent named 'Peter' occupies the place of the subject of 'ought', Peter is the normative owner of *ought to brush his teeth regularly*.

There are a variety of reasons why *IINOT* is problematic. First, treating 'ought' as a propositional operator indexed to an agent is one thing—it is a logical point about 'ought'—, but inferring from that the substantive conclusion that indexed 'ought' *is* a reliable mark of normatively owned ought is another thing. Indexation is a formal tool that serves delimiting the class of relevance of propositional operator, so the only thing that indexed 'ought' *does* show is that 'ought' is to be properly construed as ascribed to the subject of 'ought'. However, *how* we will interpret the meaning of 'indexed ought' is a matter of *philosophical* decision, and not something one can simply tell from the fact of choosing a particular logical interpretation of 'ought'. Second, even had we granted that in-

9 I originally proposed these arguments elsewhere in my *Substantive semantics*.

10 Indexing 'ought' to a subject of 'ought' (being an agent) is a logical way of representing the idea that the subject of 'ought' stands in normative relation to the action expressed in terms of the relevant proposition (action-proposition).

11 By 'relevant sort' here I mean 'ought' sentences that are most naturally interpreted as expressing personal requirements. In my preferred terminology they are 'agentive oughts' or 'truly practical oughts'.

dexed 'ought' is a reliable indicator of owned ought, *IINOT* would not be of much help since ownership is not an intrinsically normative notion. And if there are instances of ownership that are free of normative consequences for the owners, the fact of knowing who the owner of ought is—*Peter* in the discussed case—does not yet justify our jumping to the conclusion that Peter is the very person who is to be held accountable in case he violates *ought to brush his teeth*. To appreciate the worry, contrast two cases of ownership. One is evidently non-normative, whereas the other has every appearance of being normative, but is not.

Case 1: suppose that I am the owner of a juicy green sofa. I think it is evident that my being the owner of it is normatively inert. I am neither responsible for having it, nor am I to be reasonably held accountable for the consequences of having it, like the fact that the juicy green sofa is difficult to keep clean.[12] Now, case 2: consider a version of our toy example 'Peter ought to brush his teeth at least twice a day'. Assume that what the sentence in question says of Peter is true. However, from the fact that it is true of Peter that *he* ought to brush his teeth the relevant number of times a day, nothing yet follows to the effect that, necessarily, *Peter* is to be the *very* person to be held responsible for his conforming to the ought in question. Suppose that Peter is an 8 year-old boy who was given the considered instruction from his dentist. Despite the fact that Peter is the evident addressee of the 'ought to do' in question, his parents are those who ought to see to it that Peter brushes his teeth regularly. What this example suggests is that, though *ought to brush his teeth* belongs to Peter, it does not belong to him in the normatively *relevant* manner, as it is not him who is to bear normative sanctions for Peter's systematic negligence of oral hygiene.[13]

Now, why does the credibility or not of *IINOT* have any bearing on the success of Schroeder's challenge? The answer is simple: if indexed 'ought' does not

12 Some may think that this example is a bad one (thanks to the anonymous referee for pressing me to explain why the sofa example is not a case involving normative ownership. The alleged counterexamples to my sofa example are also owed to the referee). Is it not plausible to assume that being the owner of a sofa does result in some obligations, like the obligation to dispose of the sofa once one wants to trash it, or the obligation to make sure that your broken sofa will not do any harm? I do agree that having a sofa may produce some obligations like the above-mentioned ones, but I do not think that the source of these obligations comes from the fact that I am the owner of the sofa. Rather, in my understanding, they stem from some higher-order general obligations like the obligation of taking care of one's trash.

13 I discuss different sorts of normative ownership in *Substantive semantics* and "Ought', Ownership and Agentive Ought". If this example works, as I think it does, it nicely generalizes. Take any 'ought' sentence expressing a truly practical ought like 'Susan ought to run a marathon', or 'Philip ought to cook dinner', and put it in the context where the responsibility for satisfying the relevant ought is on some other agent's shoulders.

necessarily denote the normatively relevant ownership, the fact that an agent stands in the normative relation to an *unconstrained*[14] proposition does no harm since we do not need to interpret the relevant sentence's meaning the way we would interpret it were *IINOT* true. To see that the Basic Problem is a threat only if *IINOT* is true, consider again our toy example borrowed from Schroeder—'I ought that you exercise daily'. The cited sentence would be truly meaningless once we have interpreted its meaning under the guidance of *IINOT*, because then we would be forced to accept that the proposition the sentence in question expresses is that *I* (whoever the proper referent of that indexical is) *am under obligation that you* (again, whoever ithe proper referent of 'you' is) *exercise daily*. Were *IINOT* true, then the Basic Problem would be an irrefutable objection, since it is an undeniable truth that nobody is to bear the normative consequences of obligations that are not their own! But given that *IINOT* is implausible, the Basic Problem ceases to be a problem, because the allegedly odd ought-propositions are not odd, or rather they are not odd in a fatal way.

By saying that propositions like *that I ought that you exercise daily* are not odd in a 'fatal way', I mean that the content they bear is not incomprehensible, which would be the case if the odd-ought proposition in question were properly interpreted in terms of personal obligation. Nevertheless, they are still somehow bizarre. But their extraordinariness, so to speak, is not a semantic phenomenon but a syntactical one. Let me explain. If the sentence 'I ought that you exercise daily' is not a sentence about my *personal* (in the most narrow sense of the word 'personal') obligation—understood as saying that *I* am somehow to steer your body movements, so that you stick to your daily training—then the sentence makes perfect sense as an *elliptical* sentence. As an elliptical *normative* sentence it can have a variety of senses. 'I ought that you exercise daily' can be a syntactically defective sentential bearer of the proposition *that I ought to see to it that you exercise daily*. Such a proposition is easily comprehensible in the proper context, where it is part of my ought-role[15] to make sure that you exercise daily. E. g. you might be an athlete and I your coach who agreed to make sure that you will abide by the contract. To sum up this part of my considerations: I contend that the Basic Problem is a genuine problem for the interpretation of 'ought' in terms of propositional ought, *only if IINOT* is true. Since I have shown that it is not, the Basic Problem is harmless.

14 Let 'unconstrained proposition' be my name for the idea of arbitrariness of action-proposition that an agent can stand in relation to, which forms the core of Schroeder's challenge.
15 I introduced the concept of ought-role and explained its theoretical usefulness in the theory of meaning of normative 'ought' in *Substantive semantics*.

At the beginning of this section I have declared that Schroeder's Basic Problem remains in force if we accept two interpretative principles of agential 'ought' sentences with agentive content. Those principles are the *Indexation Implies Normative Ownership Thesis* (*IINOT*) and the *Ownership Implies Obligation Thesis* (*OIOT*). Above I have examined *IINOT*, and explained why it does not work. Now, let me say what is wrong with *OIOT*.

Here is what *OIOT* says: the owner of ought (being the grammatical subject of agential 'ought' sentence) is the owner of the obligation with the corresponding content. The crucial problem with *OIOT* is that it is built upon *IINOT,* and if the latter is false, its falsity obviously undermines the intelligibility of the former: if indexed 'ought' is barred credibility in the role of an indicator of owned ought, the idea according to which ownership is a mark of obligation could not even arise, as there would be no reason to think that the agent-argument place is revealing of ethically significant content.

OIOT, as I have formulated it, is untenable, not only because the principle it is built upon—*IINOT*—is false, and because ownership is not an essentially normative notion, but also because what it says is false, even in a charitable reading. In the charitable reading, *OIOT* stipulates that the normative and relevant owner of ought is the owner of obligation with the same content as the content of owned ought. Let us call this improved version of *OIOT OIOT+*. But *OIOT+* is implausible, as there is no tight conceptual or logical connection between normative ownership and obligation. This is so for two reasons. One is that normative ownership is a variety that differs with regard to strictness of normative consequences coming from not satisfying a practical ought. If you ought to buy a pair of new shoes, but failed because of laziness, it is odd to think of your failure in terms of a neglected obligation of yours, as normally getting new shoes is no obligation at all. The other is that there are instances of normative ownership that can be correctly accounted for in terms of personal obligations, yet not fulfilling them is free from normative sanctions typical for violations of one's obligations. Suppose that you ought to catch the bus and be on time for a very important professional meeting. Suppose next that not being delayed in the case at hand is not only what the norm of professional diligence requires of you but also something of more importance. Assume that at the meeting the committee discusses the employment of a new researcher and your voicing the opinion on the candidate will have a decisive influence on the dean's decision. Suppose that you believe the candidate deserves the position and you have enough time to leave home early to catch the bus. In the depicted scenario you are in capacity to catch the bus and you ought to do that, but if you did your best

and failed,[16] you would not meet your obligation 'to catch the bus and not be late for the meeting', yet no normative sanctions would follow since there are circumstances of life that prevent one from doing what one is obliged to do.

Now, how does the incredibility of *OIOT* undermine the plausibility of *BP* as applying to *RI+?* Consider again the odd 'ought' sentence 'I ought that you exercise daily'. Recall that Schroeder claims that sentences like this one are meaningless, because we cannot make sense of the idea that one agent is required to do an action that is a property of some other agent. Above, I have argued that Schroeder's commonsensical observation is to the point but *only if* we interpret the normative meaning of the sentence in question as saying that the person denoted by the indexical 'I' remains under the *obligation,*[17] very narrowly construed, that some other person denoted by the indexical 'you' will perform a particular action. However, if *OIOT* is an unreliable interpretative device, then the deviant 'ought' sentence can have different normative meaning on various occasions. For instance, the sentence 'I ought that you exercise daily' may be an elliptical sentence whose proper meaning is the one saying the following: 'that you exercise daily' is part of what I ought to do in the non-obligation sense of 'ought', because, say, I care about your health.

5 Why the Basic Problem is a Problem for Schroeder's Proposal

In section three I have argued that *IIOT* and *OIOT* are tacit assumptions underlying Schroeder's argument against interpreting truly practical 'ought' (in Schroeder's terminology 'deliberative ought') in terms of propositional ought indexed to agent, and that once you have undermined these assumptions, *BP* is toothless. In this last section, I will make a twist in the argumentative strategy in order to show that, *were BP* a genuine, irrefutable problem for the propositional interpretation of agentive ought, its destroying power would also hit the interpretation that Schroeder deems successful. This is the interpretation in which the truly practical ought is construed as a sort of property of the agent.

16 Suppose that the driver of the bus ignored your waving and did not stop at the bus stop, which was what he ought to have done. You could not have prepared an emergency plan for this occasion, as it is irrational to build one's plans starting from the assumption that people will not abide by their duties.

17 And obligations seem to be perfectly accounted for in terms of properties of agent(s).

I think that the main problem with Schroeder's Basic Problem is that his challenge raised for the propositional interpretation of 'ought' is simply unfair. The unfairness in question consists in focusing on the arbitrariness stemming from focusing on unconstrained (in my stipulated terminology) action-propositions, but ignoring the fact that treating 'ought' as a property of an agent is also arbitrary, as employing the notion of property settles in advance what properties can be sensibly owned. I call the problem in question *unfair challenge* (*UC* for short).

What I deem *unfair challenge* is raising the objection that *RI+* admits of relating agents to unconstrained action-propositions, that is, action-propositions that can only be executed by some other agents than the grammatical subject of 'ought'. I take this objection to be unfair, because if we think of the interpretation of truly practical 'ought' in terms of a relation between an agent and a property of hers to the general effect 'A is such: that A ought to φ', then our choice of the logical interpretation in question is guided by the tacit assumption that the agent can be the owner of properties over which she enjoys *personal* control. To put the issue in Schroeder's own language: it makes no sense to ask in relation to what action-property an agent stands, because she can only stand in relation to her property, whereas it makes sense to ask what action-proposition she can stand in relation to, because a proposition is not something that the agent can be in control of. But if the latter question seems to the point, while the former does not, this fact is easily explained by the trivial observation that we simply do not think of the deliberative ought in terms of actions or enterprises to be pursued by *someone other than ourselves*.

I contend that, atypical cases aside, when an agent considers what she ought to do, she selects from the options that are personally executable by her. When I ask myself a trivial question of everyday life about whether I should walk my dogs twice before going to bed, I assume that, whatever the answer I end up with, that answer is to be realized by myself since this is a practical question *for me*. Put differently, coming up with a conclusion to a genuinely practical deliberation implicates a particular method of satisfying the ought in question—authorship,[18] at least as far as is possible. Now, if we take into account that Schroeder's interpretation accounts for only one sort of truly practical oughts

18 I introduce the notion of authorship as my technical notion in *Substantive semantics*, and further discuss it in "'Ought', Ownership and Agentive Ought', and 'Ought', Agents and Ambiguity that Matters'". The notion of authorship stands for the idea of being the sole (insomuch as possible) producer of the demanded action. I stipulate that authorship is the central hallmark of the central normative ought i.e. the ought standardly (and misleadingly) referred to with the name 'deliberative ought'.

—the deliberative ones that concern *personal* actions or activities—we can reverse Schroeder's challenge and direct it at his own position by saying that it faces the same two problems that *RI+* is claimed to face. One is a version of the *overgeneration problem*, and the other a version of the *undergeneration problem* (these are Schroeder's own labels). I will add an asterisk to the problems' names to mark that the following interpretations of the respective problems apply to Schroeder's own proposal. The *overgeneration problem** runs as follows: interpreting truly practical ought in terms of a sort of agent's property reduces the deliberative ought, i.e. being the essence of truly practical ought, to one sort of deliberative ought, i.e. oughts of essentially *personal* character. Overgeneralization, as I understand it, consists in the fact of treating *any* deliberative ought as if it were necessarily *personal* in its nature, while overlooking that deliberative oughts do not make up a uniform category. Some, if not a vast majority of everyday life deliberative oughts, are personal in the sense assumed by Schroeder, but some are not. For instance, I can ponder over whether my brother should apply for a particular job position if the probability of his being successful at getting it strongly depends on my giving him a strong recommendation. The 'ought to apply for the position *X*' personally belongs to my brother as *his* deliberative question, but since his getting the job heavily impinges on my personal intervention, this is also a deliberative ought for me.

Moreover and more importantly, the result of this deliberation, when linguistically expressed, is going to take the form of the deviant 'ought' sentence of the very sort the intelligibility of which Schroeder denies. The sentence in question would be this: 'I ought that my brother will apply for the job *X*', which is an elliptical version of the full sentence saying the following: 'I ought (to make it the case by giving him professional recommendation)[19] that he will apply for the job in question'. If this example works, then it shows that Schroeder badly overgenerates by projecting a particular, very narrow interpretation of deliberative 'ought' onto the whole category of deliberative oughts, which is unfair. On the other hand though, Schroeder's own interpretation is subject to the *undergeneration problem**, which is that it can encompass only quite a narrow set of truly practical oughts, namely those that are personal oughts in the strict sense of the term 'personal', where by the strictness in question I mean the idea that those personal oughts are essentially executable only by their direct owners—e.g. ought to exercise or ought to run a marathon—leaving out quite a vast cat-

19 The content of the bracket is to be given by the context.

egory of deliberative oughts. These include ought-roles[20] like a mother's ought to educate her children, or a captain's ought to see to it that boarding is complete,[21] as well as oughts that evade an easy categorization like my ought to increase my brother's chances of getting the position.

6 Conclusion

In the paper I have set out to achieve three aims: (1) to show that Schroeder's challenge raised against the propositional interpretation of deliberative 'ought', despite its fame, is overrated to say the least; (2) to explain what assumptions give the Basic Problem the appearance of being a truly fatal objection to the sort of interpretation in question; finally (3) to demonstrate that once we give plausibility to the *BP*, the challenge turns out to be badly damaging also for the interpretation that Schroeder proposes as a promising alternative to the criticized one. I contend that I have managed to accomplish all three of them. If my arguments are not seriously flawed, they strongly support the view that the Basic Problem has received a bad fame that it does not deserve.

References

Broome, John. 2012. "Williams on *ought*". In Ulrike Hearer and Gerald Lang (eds.), *Luck, Value and Commitment: Themes from the Ethics of Bernard Williams*. Oxford University Press. 247–265.

Broome, John. 2013. *Rationality through Reasoning*. Wiley Blackwell.

Chrisman, Matthew. 2012. "'Ought and Control". *Australasian Journal of Philosophy* 90(3): 433–451.

Finlay, Stephen and Snedegar Justin. 2014. "One Ought Too Many". *Philosophy and Phenomenological Research* 86(1): 102–124.

Klimczyk, Joanna. *Substantive semantics for discourse-relevant 'ought to do'*. Book manuscript.

Klimczyk, Joanna. 2017. "'Ought', Agents and Ambiguity that Matters". *Studia Semiotyczne*. Vol. XXXI. nr 2: 113–138.

Klimczyk, Joanna. "'Ought', Ownership and Agentive Ought: Remarks on the semantic meaning of 'indexed ought". Forthcoming in *Studia Philosophiae Christianae*.

20 According to my stipulation set out in *Substantive semantics*, the concept of ought-role refers to those oughts whose most proper sentential interpretation is elliptical to the 'deviant' effect: 'A1 ought that A2 φ-s' where 'A2' denotes a different agent from 'A1'.

21 Here I assume that oughts like these, despite being deliberative oughts for the respective agents (the mother and the captain), can be satisfied by mediation of some other party or parties.

Klimczyk, Joanna. “Normative Thought”—in progress.
Merrick, Trenton. 2015. *Propositions*, OUP
Wedgwood, Ralph. 2006. “The Meaning of ‘Ought’”. In Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics*. vol. 1: 127–160. Oxford University Press.
Wedgwood Ralph. 2007. *The Nature of Normativity*, Oxford University Press.

Tomasz Puczyłowski

Gettier Cases, Warranted Assertability Maneuvers, and the Fourth Condition

Abstract: In this paper I present a warranted assertability maneuver (WAM) and three necessary criteria for its rational applications proposed by Keith DeRose. I show that these criteria are not sufficient to reject a WAM directed against Gettier's famous cases, for it satisfies all the criteria. If the WAM against Gettier's claim is correct, then Gettier's argumentation is less persuasive and conclusive than it has seemed before and it becomes questionable. If one were to insist that the Gettier cases truly showed that the traditional definition of knowledge was too broad and should not be explained in terms of generating false conversational implicatures by true assertions of Gettier's sentences, then one would have to conclude that there should be some other unidentified by DeRose criteria which were not satisfied by the WAM in Gettier case discussion. Therefore I propose the fourth standard for successful WAM: If in a context C the assertion of sentence S is unacceptable only because it generates some false conversational implicatures I_1,..., I_n, then the assertion of S in any other context C* generating false I_1,..., I_n is unacceptable to the same extent as the assertion of S in C.

1 Introduction

Presenting his version of epistemic contextualism, Keith DeRose described in detail (DeRose 2009: 81) a maneuver which in his opinion could undermine his arguments for this theory. He dubbed it the Warranted Assertability Maneuver (WAM). Other authors took up the challenge and set forth several different WAMs against the contextualist stance (e. g. Rysiew 2001, 2005, Pritchard 2005a, 2005b, Black 2005, 2008, Blaauw 2003).

However, WAM can be perceived as a general strategy which can be exploited in various debates about semantics of different types of expressions and, in fact, this line of argument is spread across different topics of modern analytical philosophy of language. It can be argued that WAMs are used not only in the still ongoing discussion concerning the contextual variability of the truth value of knowledge attributions, but the same type of reasoning can also be found in debates about the merits of direct reference theories applied to semantics of belief

Tomasz Puczyłowski, University of Warsaw (gavagai@poczta.onet.pl)

https://doi.org/10.1515/9783110702286-006

reports (Barwise & Perry 1999, Salmon 1990, 1991), or in discussions about the problem of substitution of co-referring proper names in some simple sentences (Barber 2000). One can also find WAM-like arguments in philosophical debates about the problem of the alleged meaningfulness of sentences with fictional names (Adams 2002, Adams & Dietrich 2004, Adams & Fuller 2007, Adams & Stecker 1994).

Interestingly, DeRose proposed three criteria to evaluate the merits of a WAM. I am going to argue that they are not sufficient for the decisiveness of WAM-like arguments, thus there is at least another necessary condition which must be met by a WAM for its applications to be rational or convincing. The main aim of this paper is to formulate the fourth necessary criteria for a successful WAM. In order to do that, I will present a WAM to the famous Gettier cases (Gettier 1963). Its purpose is to show that an assertion of a knowledge attribution (in the form of a Gettier sentence) may be true if it satisfies the conditions of the traditional definition of knowledge, and yet it may seem to be unacceptable and unassertable only because its assertion would generate a false generalized conversational implicature (GCI). It is believed that assertions which convey false GCIs often give an impression of being false, therefore an assertion of the Gettier sentence is expected to give the appearance of falsity. However, I am going to show that the WAM against Gettier's claim (i. e. the claim that the traditional definition of knowledge is false) satisfies all three criteria proposed by DeRose, and thus if it is to be discharged, there must be another not met. In order to discover and explicitly formulate it, I will discuss a case which is similar to Gettier's one in terms of implicatures generated by the assertion of the Gettier sentence and their truth value; yet in this case the sentence will be apparently warrantably assertible and true. Because the scope of applications of WAM is wide and extends to other philosophical problems and topics, there is a pressing need from a metaphilosophical perspective for a complete list of conditions for rational and successful WAM; I believe the conditions to be of great importance, as they are essential tools for evaluating frequent WAM-like arguments spread across modern analytical philosophy of language and epistemology.

2 Warranted Assertability Maneuver

The Warranted Assertability Maneuver is a specific response to a common argument (ARG) against a semantic theory that in a general, if not oversimplified, form, can be put as follows:

(ARG)

1) That a sentence S is true in a context C is a consequence of a theory T (alternatively, I will be using the phrase: According to a theory T, a sentence S should be true in a context C);
2) However, S seems to be false in C;
3) S in C seems to be to false, because assertion of S in C is not warranted in C;
4) If a sentence is not warrantably assertible in a context and there is no other plausible explanation for that, it is just false in the circumstances;
5) There is no other plausible explanation why S in C seems false;

Therefore

C) T is false.

The WAM is directed against premise 5), i.e. it attempts to show that there are other reasons responsible for the apparent falsehood of an assertion. In its shortest form, it claims that the reason why an assertion of a sentence seems to be false in a given context is that some false generalized conversational implicatures are generated by the assertion of the sentence in that context. That claim, together with a belief that we tend to confuse the semantic content of a sentence with the conversational implicatures of its assertion, is to explain why a given sentence, even if true, could seem—at least to some of us—to be false. So, with reference to (ARG), WAM can be put in the following form:

(WAM)

There is a plausible explanation why S seems false in C, i.e.:

1) An assertion of S in C conversationally implies a proposition P;
2) If an assertion of a sentence conversationally implies in a context a false proposition, the sentence is not warrantably assertible in that context and may seem to be false;

Therefore

A sentence S is not warrantably assertible in C, and S seems to be false in C.

Unfortunately, it is not clear what precisely is meant by the notion of warranted assertibility. However, it is safe to say that if an assertion of a sentence does not satisfy some conditions, it is not warranted. What sort of conditions? I do not dare to specify them all, but some uncontroversial candidates could be put forward. For instance, if a sentence is syntactically not well-formed, its assertion is

probably not warranted. Even if we can understand in some specific context what one means by uttering

(1) Let's to the bar together go,

it is not warranted, at least not by English grammar. Also, if one answers the question 'what is the capital of France?', and by wild guessing sincerely insists that it is Berlin, one's assertion could seem unwarranted, if one truly lacks any evidence for that claim. If someone says that some widows haven't had any husbands, one definitely breaks some semantic rules of English, as well as when one says without irony that one's mother brother is not one's uncle. Probably, if one says that Barbara is older than Ted, but immediately adds that Ted is not younger than Barbara, at least one of those assertions is not warranted. If it rains somewhere, but one who witnesses that rainy event says that it is not raining, the assertion is not warranted. Finally, if one knows that Peter is at the library every day and is asked when Peter is present there, but decides to answer by saying that he is there on Mondays or Fridays, then probably one's assertion would count as unwarranted, at least partially, for it implies something plainly false, i. e. that one does not know precisely on what days Peter is present at the library. If the given examples are intuitive and correct instances of unwarranted assertions, then one can come to the plausible conclusion that an assertion of a sentence is not warranted if the sentence does not comply with some grammatical, semantic or pragmatic norms or principles; an assertion is unwarranted if it is false or implies something false, or if it is ungrammatical, or when one does not have any warrant for the conveyed information. However, other norms concerning politeness etc. do not seem to apply when considering the warranted assertability of a sentence. In other words, even if there were strict enough standards for being handsome and ugly, someone's assertion that Peter is ugly would be warranted, though rude, were it truly the case, but, on the other hand, it would be unwarranted in case Peter were undeniably handsome.

Although the problem of the definition of warranted assertability is important and interesting, it is not crucial for further discussion. In fact, (ARG) could take a simpler form, such as the following:

(ARG*)

1) According to a theory T, a sentence S should be true in a context C;
2) However, S seems to be false in C;
3) If a sentence seems to be false in C and there is not any other plausible explanation for that, it is just false in the circumstances;
4) There is no other plausible explanation why S in C seems false;

Therefore

C) T is false.

(WAM) adjusted to (ARG*) needs to explain why an assertion of a sentence seems to be false in a given context, and it achieves that end by pointing to an almost universal principle. According to that principle, when any assertion gives rise to a false conversational implicature, then either the sentence or its semantic propositional content seem to be false.

It is obvious that an unrestricted WAM could be used to defend undoubtedly false semantic theories, therefore DeRose (1999: 201–203, 2002: 172–176) argues that (WAM) is worth considering in defense of a semantic theory T only if WAM satisfies the following three conditions:

(CRITERIA)

WAM is successful only if

1. it needs to explain only an appearance of falsity—in particular, not only a given sentence must appear to be false in one and the same context, but also its negations must seem to be false;
2. it can provide the explanation by appealing only to the generation of a false implicature (or false implicatures);
3. it appeals to general rules of conversation in explaining why the apparently false assertion is unwarranted.

It is not the subject of the paper to argue against DeRose's proposal, nor is it in my interests to analyze and discuss here the rationale for the proposed conditions given by DeRose. Nevertheless, it has to be mentioned that some argue that he has not provided a correct set of conditions for a successful WAM (e.g. Black 2008), but I will not address that issue here.

However, when it comes to the first condition, it is worth mentioning that it is not clear how to understand it. According to the first of its interpretations, for all competent users of a given language both a sentence S and its apparent negation must seem false in the same context. According to the second, it suffices that there are competent users who consider S to be false (in a given context) and that there are other users who (in the same context) take apparent negation of S to be false. We assume, though, that the condition, especially in the second meaning, is satisfied by the WAM against Gettier's claim (i.e. the claim that the traditional definition of knowledge is false; from now on the WAM against the claim will be dubbed 'WAM against Gettier')). The assumption stems from the

fact that there is some empirical data which indicates, against Gettier's and his followers' intuitions, that some competent agents tend to attribute knowledge to the Gettier's subjects.[1] Even if some of the results of empirical studies (Nisbett *et al.* 2001, Weinberg *et al.* 2001) don't directly support the observation that some people don't share the intuitions of some epistemologists with regard to the Gettier cases, there are some interesting studies (Starmans & Friedman 2012, Turri 2013) that seem to support that conclusion or at least undermine the claim that the WAM against Gettier should be rejected on the grounds of non-fulfillment of the first condition of a successful WAM. So, from now on, I will focus on showing that the other two conditions are satisfied by the WAM formulated against Gettier cases.

Now, I maintain that DeRose has not provided a complete set of conditions for a successful WAM. For if it were a complete and correct set, the following WAM against Gettier would be successful, and the long-lasting discussion of the Gettier cases would be resolved. However, one can think of the Gettier problem as substantial, not just a matter of pragmatics (Bogusławski 2002) and thus seek an overlooked reason to reject any WAM directed against Gettier's point.

3 WAM and the Gettier Problem

Before I present the WAM against Gettier, let us introduce some useful abbreviations. For the sake of convenience, let 'Bxp' stand for 'x believes that p', 'Tr(p)' for 'it is the case that p', 'Kxp' for 'x knows that p', 'JBxp' for 'x's belief that p is justified'; let's read 'A $\models$ B' as 'A entails B', and let propositional variables represent sentences of a given language expressing propositions (I assume that different variables represent different sentences expressing different propositions).

In general, the Gettier case is a scenario in which a subject x actively considers at least three propositions and the following holds for a given subject x:

1. $\sim Tr(p_1) \wedge Bxp_1 \wedge JBxp_1$
2. $Tr(p_2) \wedge \sim Bxp_2 \wedge \sim JBxp_2$
3. $p_1 \models p_3$
4. $p_2 \models p_3$
5. non $(p_3 \models p_1)$
6. non$(p_3 \models p_2)$

1 See e.g. Shope 1983, Sosa 2007, Zagzebski 1996.

If it can be assumed that

7. $\forall\alpha\ \forall\beta\ ((Bx\alpha \wedge \alpha \models \beta) \Rightarrow Bx\beta)$ and
8. $\forall\alpha\ \forall\beta\ ((JBx\alpha \wedge \alpha \models \beta) \Rightarrow Bx\beta)$,

then

9. $Tr(p_3) \wedge Bxp_3 \wedge JBxp_3$.

If we adopt the so-called traditional definition of knowledge, according to which x knows that p iff x's belief that p is true and justified, then we can conclude that a subject from the Gettier case knows that p_3. But, according to Gettier and a vast number of epistemologists who have followed his judgement in this respect, the conclusion is unlikely or plainly false.

For further reasons, and to shed some light on the logical structure of the Gettier cases, I would like to present the first Gettier case in the form of a matrix:

Table 1: the first Gettier case.

Proposition:	Is it true?	Does Smith believe that?	Is Smith's belief justified?	According to the traditional definition of knowledge, does Smith know that?	According to intuition/common sense, does Smith know that?
(G1) Jones has a Ford	No	Yes	Yes	No	No
(G2) Brown is in Barcelona	Yes	No	No	No	No
(G3) Jones has a Ford or Brown is in Barcelona	Yes	Yes	Yes	Yes	No

A similar matrix could be presented for the second case, but for the sake of brevity I leave it to the reader.

Putting the Gettier case in the form of (ARG*), we get:

(ARG*Gettier)

1. According to the traditional definition of knowledge, a Gettier sentence i.e. '*Smith knows that Jones has a Ford or Brown is in Barcelona*' is to be true in a described context;
2. However, the Gettier sentence seems to be false in that context;
3. If a sentence seems to be false in a given context and there are no other circumstances which could explain why it seems so, the sentence is just false

4. There are no other circumstances which could explain why the Gettier sentence seems to be false in the described context

Therefore

5. The traditional definition of knowledge is false.

The WAM against (ARG*Gettier) that I am going to propose will undermine premise 4., namely it will be argued that an assertion of the Gettier sentence generates a false conversational implicature (in the described context).

In order to do that, I will refer to a principle which seems to be a well-established law of pragmatics (see Levinson 1983: 136, Lenzen 2003):[2]

(LAW)
If (i) S is not more lexically complex, elaborated etc. than S*, and
(ii) S expresses a proposition p and S* expresses a proposition p* and
(iii) S entails S*, and
(iv) both p and p* are equally relevant in a context C, but
(v) a speaker X asserts in C only S*, then
X conversationally implies in C that she doesn't believe p.

The idea behind this principle is based on Grice's maxim of quantity. If a speaker chooses to assert a weaker proposition instead of a contextually relevant one that is stronger, then if she chooses so not because of the need to be brief, elegant, intelligible etc., she allows us to conclude that she doesn't believe the stronger one or has no evidence for it whatsoever.

First, let's observe that on the ground of epistemic logic one can maintain that a proposition expressed by

(G1) Smith knows that Jones has a Ford

is stronger than (i.e. entails) the proposition expressed by the Gettier sentence

(G3) Smith knows that Jones has a Ford or Brown is in Barcelona.

2 Lenzen says (2003): "The basic idea of Grice's maxim of quantity can now be rendered more precise in the following way. If the speaker a has the choice between two assertions p and q, where the former is relevantly more informative than the latter, then a is conversationally obliged to make the more informative assertion p provided this is not in conflict with the maxim of quality. In other words, a is conversationally allowed to make the less informative assertion q only if a is not certain that p."

It is so, because it is a tautology of classic propositional logic that $p \Rightarrow (p \vee q)$, which by the necessity rule Nec allows the inference $Kx(p \Rightarrow (p \vee q))$, which by axiom (K) of modal logic adjusted to epistemic logic leads to $Kxp \Rightarrow Kx(p \vee q)$. Also, it is evident that (G1) is less lexically complex than (G3), if we agree that p is less lexically complex than $(p \vee q)$, i.e. that any proposition is less lexically complex than its alternative with any other proposition.

It should now be clear that if we take (LAW) and all the observations for granted, then it follows that anyone who asserts (G3) in a context where the state of Smith's knowledge is relevant, conversationally implies that she doesn't believe (and therefore doesn't know) the proposition expressed by the sentence (G1). In short, if a speaker says (in the context) that Smith knows that Jones has a Ford or Brown is in Barcelona, her sentence conversationally implies that she doesn't know whether Smith knows that Jones has a Ford. By saying that, she also conversationally implies that she doesn't know whether Smith knows that Brown is in Barcelona. It is not hard to notice that these implicatures (i.e. propositions implied by a speaker who says that (G3)) are false in the described context. Anyone who knows that context, described so vividly by Gettier, already knows that Smith doesn't know anything relevant! Clearly, he doesn't know that Jones has a Ford, and he doesn't know that Brown is in Barcelona. And we all know it. But by saying (G3), we would be conversationally implying otherwise. We would imply that we don't know whether Smith knows that Jones has a Ford. So that is the reason we are reluctant to express the Gettier sentence, and therefore we reject the proposition that Smith knows that (G3). Therefore, one can conclude that we are wrongly hesitant to attribute knowledge to Smith, and in consequence we feel compelled to reject the definition of knowledge, because our assertion of a true Gettier sentence would conversationally imply something plainly false about our propositional attitudes in the context.

Of course, one can apply the same kind of reasoning to the second Gettier case, the only difference is that one needs to refer to the existential generalization law of first-order logic instead of the tautology of propositional calculus.

4 Tests for Conversational Implicature

It is widely accepted that conversational implicatures have some distinctive features that differentiate them from semantic consequences and conventional implicatures. The former must be cancellable, calculable, non-conventional, and (except those generated by the maxim of manner) undetachable, the latter are non-cancellable, conventional, and in the case of conventional implicatures, detachable. Therefore, before I move forward, I should verify the claim that an as-

sertion of (G3) conversationally implies that a speaker doesn't believe (and therefore doesn't know) the proposition expressed by (G1) by establishing if alleged implicatures satisfy the conditions imposed by H. P. Grice on conversational implicatures.

Let us begin with the calculability of conversational implicatures. This feature demands that for every supposed implicature there must be given an argument, which—by exploiting what was said in a circumstance, some further features of the context, and the Cooperative Principle along with the conversational maxims—explains how a speaker who says something conveys a given proposition. The proposition in WAM against Gettier is clearly calculable from the conversational maxims, if it can be taken for granted that it is calculable that the assertion of a disjunction conversationally implies that a speaker doesn't know which of the disjuncts is true. Let's turn to Levinson, who univocally states (Levinson 1983: 143): "(...) if one knows that p, one does not co-operatively convey that by stating p or q; the use of disjunction rather conveys that one has grounds for believing one or the other disjunct but does not know which."

If anything is taken for granted in a theory of conversational implicatures, it is that an assertion of the form $(p \vee q)$ conversationally implies that a speaker doesn't know that p and doesn't know that q. This conviction is based on the reasonability of (LAW), and there are good reasons to accept (LAW) as well as for the acceptance of $Kxp \Rightarrow Kx(p \vee q)$. These reasons are elegant and convincing: Grice's conception of rational conversation on the one hand, and the soundness of modal and epistemic logic on the other. (LAW) and some laws of epistemic logic are everything that is necessary and sufficient to calculate how an assertion of the Gettier sentence conversationally implies that a speaker doesn't know whether Smith knows that Jones owns a Ford.

Now let's turn to the rest of the necessary conditions for a proposition to be conversational implicature of a speaker's utterance. However, before I discuss whether proposed propositions could be conversational implicatures of someone's assertion of the Gettier sentence, I should make clear how these defining features of the implicatures are to be understood. First, I propose the following definition of non-conventionality:

(NONCONV)

If an utterance of a sentence S conversationally implies in a context C a proposition p, then the proposition is a nonconventional part of the content conveyed by an act of assertion iff there is a different context C* in which utterance of S does not conversationally imply p.

In order to clarify and fully grasp (NONCONV), let's look at an example. Imagine you are asked in the evening if you are hungry, and you reply 'I've had breakfast'. Your assertion would truly conversationally imply that you (i.e. a speaker) are hungry, if there is a different context in which the assertion of the same sentence would not imply the same proposition. And, of course, there is such a context: if your reply is the same but takes place in the morning, then you clearly do not imply that you are hungry, for it is evident that you suggest just the opposite.

To prove that the propositions conversationally implied by the assertion of (G3) are a non-conventional part of the content of the assertion, it suffices to realize that, in the context where (G3) is a direct answer to the question 'Does Smith know that Jones has a Ford or Brown is in Barcelona?', the assertion of (G3) does not convey the proposition implied by (G3) in the Gettier context.

This observation exploits a test for quantity and relevance implicatures, according to which:

(TEST)

If an assertion of a sentence S in a context C conversationally implies a proposition p, then S is not a direct answer to a question 'Is/Are/Am/Do/Does/Did/Will/Would S*?' (where S* is a proper form of erotetic sentence transformed from a declarative sentence S).

The reason behind (TEST) is that these are the questions that define what the aim of the exchange of words in which we participate is. Also, the other role of questions is to specify to others what is relevant information and how much information is required in the context. Therefore, if one answered 'Some lads were happy' to the question 'Were some lads happy?', there would not be an implicature that, according to the speaker, not all lads were happy. However, that proposition would be implied if the question were 'Were all lads happy' or 'How many lads were happy?' or, perhaps, 'Who was happy?'.

Non-detachability is a property of conversational implicature by which I propose to understand:

(NONDET)

If an utterance of a sentence S in a context C conversationally implies a proposition p, then the proposition is a non-detachable part of the content of the utterance iff an utterance in the C of any sentence synonymous to S conversationally implies the proposition.

For example, consider an assertion of 'John is taller than Ted', being uttered as a reply to a question 'Who is the better basketball player: John or Ted?'. If it conversationally implies (in a given context) that John is a better basketball player than Ted, then the same conversational implicature will be generated by an assertion of a different, yet synonymous sentence: 'Ted is smaller than John'.

If someone's assertion 'John is in the garden or in the kitchen' conversationally implies (in a given context) something (for example: 'I don't know where precisely John is: in the kitchen or in the garden'), then the same would be conversationally implied if a logically equivalent sentence 'If John is not in the garden, then he is in the kitchen' was asserted.

I claim that if in the Gettier case we were to estimate the truthfulness or acceptability of a logical equivalent to the original Gettier sentence like 'Smith knows that if John does not have a Ford, then Brown is in Barcelona', the result would be the same. For an assertion of the sentence 'Smith has a Ford or Brown is in Barcelona' conversationally implies (in a given context) the same as the assertion of 'If Smith doesn't have a Ford, then Brown is in Barcelona'.

Now, let's turn to the last property of conversational implicatures—cancellability. Usually, it is defined (or rather described) in a way that is English-oriented, i.e. by using phrases specific to English, but I would like to propose a definition which is general enough to apply also to other languages.

(CANC)

If an utterance of a sentence S in a context C conversationally implies a proposition p, then the proposition is cancellable iff there is a sentence S* the admissible assertion of which in C does not conversationally imply p, but (i) S is entailed by S*, and (ii) S is a proper part of S*.[3]

The alleged conversational implicatures of an assertion of the Gettier sentence are clearly cancellable. It suffices to notice that there is nothing absurd or wrong in saying:

(2) *Smith knows that Jones has a Ford or Brown is in Barcelona, in fact, I know which of the two he knows*,

(3) *Smith knows that Jones has a Ford or Brown is in Barcelona, because I know that he knows the first (i.e that Jones has a Ford)*,

3 There is an on-going, very interesting debate as to how to define this feature (see e.g. Blome-Tillmann 2008, Weiner 2006)

(4) *Smith knows that Jones has a Ford or Brown is in Barcelona and I think that he knows whether Jones has a Ford.*
(5) *Smith knows that Jones has a Ford or Brown is in Barcelona, but I don't want to imply that I don't know which he knows.*

or even the quite elaborated and somewhat mysterious (if we don't take for granted that the traditional definition of knowledge is false)

(6) *Smith knows that Jones has a Ford or Brown is in Barcelona, although it is not the case that he knows that Jones has a Ford, and it is not the case that he knows that Brown is in Barcelona.*

All the sentences contain (as their proper parts) (G3), and (G3) is entailed by every one of them, but their assertions clearly would not imply that a speaker does not believe the proposition expressed by (G1).

I don't wish to claim that the described warranted assertability maneuver against Gettier's stance is convincing or decisive. It is not, as I will show in the next part. My point is that the WAM satisfies all three criteria proposed by DeRose, therefore, in order to reject it, we need to find a condition overlooked by DeRose and others. Let's turn to the fourth criterion for a successful WAM.

5 The Fourth Condition

Before I formulate the fourth necessary condition for a successful WAM, let us picture a scenario similar to the one originally described by Gettier. In the new circumstances everything is just like in Gettier's, but this time Jones is the owner of a Ford, therefore the proposition that he has a Ford is true. Consequently, the status of Smith's epistemic position is radically changed. In the new scenario, he truly knows that Jones has a Ford. Of course, knowing that, he also knows that Jones has a Ford or Brown is in Barcelona. This time, however, I believe there is no divergence between our intuition to attribute knowledge to Smith and the verdict given by the traditional definition of knowledge.

I present vital data for this scenario in the following table:

Table 2: The modified Gettier case

Proposition:	Is it true?	Does Smith believe that?	Is Smith's belief justified?	According to the traditional definition of knowledge, does Smith know that?	According to intuition/common sense, does Smith know that?
(G1) Jones has a Ford	Yes	Yes	Yes	Yes	Yes
(G2) Brown is in Barcelona	Yes	No	No	No	No
(G3) Jones has a Ford or Brown is in Barcelona	Yes	Yes	Yes	Yes	Yes

What is interesting and needs to be emphasized is that an assertion of (G3), in the new scenario, would conversationally imply the same proposition that the speaker does not know (or believe) that (G1). This conversational implicature of (G3) would be in the new circumstance false, just like in the original. If the falsehood of the conversationally implied propositions was to explain the discrepancy between the verdict of the traditional definition of knowledge and our intuitions concerning Smith's epistemic status, one would expect that in the new scenario a claim that Smith knows that Jones has a Ford or Brown is in Barcelona would seem false. But, clearly, it does not. It looks like, in the new circumstances, we are ready to assert without much hesitation that Smith knows that Jones has a Ford or Brown is in Barcelona. Even if there is some sort of a hunch that this assertion is not completely acceptable, the sensation is far from the one which is evoked in the original Gettier case, if only one tries to assert the Gettier sentence.

Now I am ready to explain why a WAM for the Gettier problem is not very convincing. I believe that the fourth condition overlooked by DeRose could be stated as follows:

(THE FOURTH)

If an assertion of a sentence S in a context C is not to be acceptable (assertible) due only to conversationally implying false conversational implicatures p_1,..., p_n, then an assertion of S in any other context C* in which false p_1,..., p_n are generated should not be acceptable (assertible) (to the same extent as in C).

The WAM against Gettier does not satisfy that condition. Of course, there may be some false proposition, other than that discussed above, which is conversation-

ally implied by the assertion of the Gettier sentence in the considered context. If there is one which is generated in the original Gettier case, and if it is false in the same context, and if it is not generated in the new scenario, or if it is generated, but it is not false in it, then the WAM I presented here against Gettier is not proven unsuccessful. Yet, until proven otherwise, I do not see any such proposition; so, in conclusion, I maintain that the WAM against Gettier is inconclusive, for it does not satisfy the criterion that I set forth here.

References

Adams, Frederick. 2002. "Names that Name Nothing". In: Christian Kanzian, Josef Quitterer, Edmund Runggaldier (eds.), *Papers of the 25th International Wittgenstein Symposium*. Kirchberg: Austria, 8–10.

Adams, Frederick and Dietrich Laura. 2004. "What's in a(n Empty) Name?". *Pacific Philosophical Quarterly* 85: 125–148.

Adams, Frederick and Fuller Gary. 2007. "Empty Names and Pragmatic Implicatures". *Canadian Journal of Philosophy* 37: 449–462.

Adams, Frederick and Stecker Robert. 1994. "Vacuous Singular Terms". *Mind and Language* 9: 387–401.

Barber, Alex. 2000. "A Pragmatic Treatment of Simple Sentences". *Analysis* 60: 300–308.

Barwise, Jon and Perry John. 1999. *Situations and Attitudes*. Stanford CA: CSLI Publications.

Blaauw, Martijn. 2003. "WAMming Away at Contextualism". *SATS* 4: 88–97.

Black, Tim. 2005. "Classic Invariantism, Relevance, and Warranted Assertability Manœuvers". *The Philosophical Quarterly* 55: 328–336.

Black, Tim. 2008. "Warranted-assertability Defense of a Moorean Response to Skepticism". *Acta Analytica* 23: 187–205.

Blome-Tillmann, Michael. 2008. "Conversational Implicature and the Cancellability Test". *Analysis* 68(298): 156–160.

Bogusławski, Andrzej. 2002. "There is No Getting Round Gettier". *Journal of Pragmatics* 34 (8): 921–937.

DeRose, Keith. 1999. "Contextualism: An Explanation and Defense". In: John Greco and Ernest Sosa, (eds.), *The Blackwell Guide to Epistemology*. Oxford UK: Blackwell Publisher, 196–203.

DeRose, Keith. 2002. "Assertion, Knowledge, and Context". *Philosophical Review* 111 (2): 167–203.

DeRose, Keith. 2009. *The Case for Contextualism: Knowledge, Skepticism and Context, Vol. 1*. Oxford: Oxford University Pres

Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?". *Analysis* 23(6): 121–123.

Grice, Herbert Paul. 1989a. "Further Notes on Logic and Conversation". In: H. P. Grice *Studies in the Way of Words. Cambridge*, (MA): Harvard University Press. 41–57.

Grice, Herbert Paul. 1989b. "Logic and Conversation". In: Herbert Paul Grice, *Studies in the Way of Words, Cambridge*. (MA): Harvard University Press.

Jackson, Frank. 2011. "On Gettier Holdouts". *Mind & Language* 26(4): 468–481.

Lenzen, Wolfgang. 2003. "Knowledge, Belief, and Subjective Probability: Outlines of a Unified System of Epistemic/Doxastic Logic". In: Vincent Hendricks, Klaus Frovin Jørgensen, Stig Andur Pedersen, (eds) *Knowledge Contributors. Synthese Library (Studies In Epistemology Logic, Methodology, and Philosophy of Science, vol 322.* Dordrecht: Springer

Levinson, Stephen C. 1983. *Pragmatics.* Cambridge UK: Cambridge University Press

Lycan, William G. 2006. "The Gettier Problem Problem". In: Stephen Hetherington (Ed.). *Epistemology futures.* Oxford: Oxford University Press. 148 – 168.

Nagel, Jennifer, San Juan Valerie and Mar Raymond A. 2013. "Lay Denial of Knowledge for Justified True Beliefs". *Cognition* http://dx.doi.org/ 10.1016/j.cognition.2013.02.008.

Nisbett, Richard E., Peng Kaiping, Choi Incheol and Norenzayan Ara. 2001. "Culture and Systems of Thought: Holistic Versus Analytic Cognition". *Psychological Review* 108(2): 291 – 310. http://dx.doi.org/10.1037/0033 – 295X.108.2.291.

Pritchard, Duncan 2005a. "Contextualism, Skepticism, and Warranted Assertibility Maneuvres". In Joseph Klein-Campbell, Michael O'Rourke, Harry Silverstein (eds). *Knowledge and Skepticism.* MIT Press

Pritchard, Duncan. 2005b. "Neo-Mooreanism versus Contextualism". *Grazer Philosophische Studie*n 67. 2043.

Rysiew, Patrick. 2001. "Contex-Sensivity of Knowledge Attributions". *Nous* 35:4. 477 – 514

Rysiew, Patrick. 2005. "Contesting Contextualism". *Grazer Philosophische Studien* 69: 44 – 62.

Salmon, Nathan. 1990. "A Millian Heir Rejects the Wages of Sinn". In: Anthony C. Anderson, Joseph Owens (eds.). *Propositional Attitudes. The Role of Content in Logic, Language, and Mind.* Stanford: Center for the Study of Language and Information. 215 – 247.

Salmon, Nathan. 1991. *Frege's Puzzle (Second Edition).* Atascadero CA: Ridgeview.

Shope, Robert K. 1983. *The Analysis of Knowing*, Princeton: Princeton University Press

Sosa, Ernest. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, volume I. Oxford: Oxford University Pres

Starmans, Christina and Friedman Ori. 2012. "The Folk Conception of Knowledge". *Cognition* 124.3: 272 – 283.

Turri, John. 2011. "Manifest Failure: The Gettier Problem Solved". *Philosophers' Imprint* 11 (8): 1 – 11.

Turri, John. 2013. "A Conspicuous Art: Putting Gettier to the Test". *Philosophers' Imprint* 13(10): 1 – 16.

Weinberg, Jonathan, Nichols Shaun and Stich Stephen. 2001. "Normativity and epistemic intuitions". *Philosophical topics* 29: 429 – 460.

Weiner, Matthew. 2006. "Are All Conversational Implicatures Cancellable?". *Analysis* 66(2): 127 – 130.

Zagzebski, Linda Trinkaus. 1996. *Virtues of the Mind: an Inquiry Into the Nature of Virtue and the Ethical Foundations of Knowledge.* Cambridge: Cambridge University Press.

Andrew Sneddon

Self vs Other? Social Cognition, Extended Minds, and Self-Rule

Abstract. Humans are individuals qua objects, organisms and, putatively, minds. We are also social animals. We tend to value self-rule—i.e., the possession and exercise of the capacity or capacities that allow individuals to govern their lives. However, our sociality can call the possibility and value of such autonomy into question. The more we seem to be social animals, the less we seem to be capable of running our own lives. Empirical psychology has revealed surprising details about the extent to which our minds are subject to social influence. Such influence is sometimes taken as a threat to self-rule. The debate over the Extended Mind Hypothesis (EMH) might seem to exacerbate the social threat to self-rule. If minds are spread across individuals, perhaps it no longer makes sense to speak of individual selves who may or may not rule themselves. I argue that no general threat to self-rule stems from human sociality. Further, the adoption of EMH is consistent with extensive attribution of psychological states to individuals, thus vouchsafing individual selves and autonomy.

1 Introduction

Humans are individuals *qua* objects, organisms and, putatively, minds. We are also social animals. We tend to value self-rule, i.e. the possession and exercise of the capacity or capacities that allow individuals to govern their lives. However, our sociality can call the possibility and value of such autonomy into question. The more we seem to be social animals, the less we seem to be capable of running our own lives.

So much is clear without delving into philosophical debate about the metaphysics of cognition. However, the debate over the Extended Mind Hypothesis (EMH) might seem to exacerbate the social threat to self-rule. So-called individualists defend the idea that the mind must be, strictly speaking, bodily bounded; so-called externalists, proponents of EMH, deny this. There are various ways of making the case that cognition is extended. One is in terms of social relations. The general idea is that particular cognitive processes should be thought of as taking place between people, rather than within them. Once we start to think

Andrew Sneddon, University of Ottawa (Andrew.Sneddon@uottawa.ca)

https://doi.org/10.1515/9783110702286-007

this way, one might well wonder whether we are mistaken to think of people as discreet selves. If we are mistaken, then we are also mistaken about the possibility of autonomy: no selves, no self-rule.

I shall argue that the worries about autonomy that might be generated by empirical and theoretical work on the social dependence of cognition are misplaced. Neither the empirical discovery of the extent to which we are social animals nor a more theoretical account of the metaphysics of minds which allows cognition to be spread among people should shake our view of ourselves as capable of controlling our own lives. Autonomy is both possible and, in principle, valuable for social animals such as ourselves.

2 Intuitive Challenges to Self-Rule from Social Cognition

Humans are deeply social animals. On-going empirical study of human cognition has revealed that we have specific mechanisms for social cognition and that our thinking in general is subject to social influence. Arguably it's not just that our minds exhibit social sensitivity, it's that they are made for this. What should we make of this view of the mind? The social aspects of individual cognition provide the starting point for what I will call 'intuitive' challenges to autonomy. By 'intuitive' I mean that these challenges are a first thought or a gut feeling rather than well-worked out positions. Nevertheless, such starting points are important for a variety of reasons. These are the ideas that people work with before they have reflected on matters, which means that they can be very powerful. After all, people might make decisions about action and policy using such ideas without ever really reflecting on them. Moreover, reflection and argumentation often take the form of attempts at vindicating our first thoughts, rather than of genuine efforts to assess whether our intuitions are true. Either way, the force and staying power of such rather inchoate ideas makes them worth our attention.

Several intuitive challenges to autonomy find their foundation in our social minds. Here are the broad strokes. Autonomy is self-rule. However, the emerging evidence about social aspects of human psychology problematizes this notion. Our thoughts and actions do not issue from us alone. They are the product of complex interaction between us and other people, not to mention other features of the world. The worry is this: while we might well be 'selves' in some important sense, we are not self-ruling.

Here are some examples of the empirical bases of this worry:

a) Social psychology in general is a good place to look for evidence of the influence of other people on our minds. Such evidence—perhaps surprisingly, perhaps not—is found for the values that we typically take as defining our deepest commitments. Anthropologists and psychologists have collected information about links between values and their psychological roots (Haidt *et al.* 1993, Shweder *et al.* 1997, Rozin *et al.* 1999, Haidt & Bjorklund 2008). This has been influentially put in terms of a 'triad' of values and emotions. However, subsequent work has expanded the triad. The most influential statement of this tradition now comes from Jonathan Haidt, who contends that there are six 'moral foundations', although these domains overlap to some extent with the original three (Haidt & Joseph 2007, Iyer *et al.* 2012). I shall focus on the original triadic formulation.

The values most familiar to western philosophers fall into the class of values that Richard Shweder and colleagues name 'autonomy.' This domain "relies on regulative concepts such as harm, rights, and justice... and aims to protect the discretionary choice of 'individuals' and to promote the exercise of individual will in the pursuit of personal preferences" (Shweder et. 1997: 138). It is linked to the emotion of anger.

On the basis of an analysis of moral discourse of residents of Bhubaneswar, India, Shweder and colleagues add two more classes of values. First there is the class called 'divinity,' which "relies on regulative concepts such as sacred order, natural order, tradition, sanctity, sin, and pollution" (*ibid:* 138). This is related to the emotion of disgust. The other class of values is 'community,' which "relies on regulative concepts such as duty, hierarchy, interdependency, and souls.... It aims to protect the moral integrity of the various stations or roles that constitute a 'society' or 'community', where a 'society' or 'community' is conceived of as a corporate entity with an identity, standing, history, and reputation of its own" (*ibid*). Schweder and colleagues link this to contempt.

Two things are worth emphasizing about this body of work. First, the roots of our values are biological. However, and second, the ways in which we think about values—for instance, which ones we tend to prioritize—are deeply affected by social context. People in one place and time will tend to emphasize, e. g., the autonomy values, whereas other people will grow up in a context that emphasizes one of the other wings of the triad. The psychological roots are present in all of us, but they are developed in different mixes depending on just who we grow up and live with. There are, of course, individual differences in the ways in which such influence works, just as there are individual differences in biology. This does not call into question the fact that even such a fundamental feature of our lives as our deepest commitments is subject to social influence from square two (after biology).

b) Not just our ideas about values are subject to social influence. Social psychology has famously documented the effects of social interaction on the production of action. The well-known debate between personality psychologists and so-called 'situationists' provides lots of examples. Very roughly, personality psychologists argue that variation in behavior between individuals is due to variation in certain sorts of psychological traits possessed by those individuals. By contrast, and just as roughly, situationist social psychology argues that variation in behavior is due much more to differences in situations than we are inclined to think. This debate was sparked by Walter Mischel's review of the literature on personality and action production in *Personality and Assessment* (1968). At the beginning of the twentieth century situationist tradition stands *Studies in Deceit* by Hugh Hartshorne and Mark May (1928). Hartshorne and May performed a long-term study of deceit involving thousands of children in classroom settings. They used a variety of tests to assess their subjects for deception and honesty in various forms, such as cheating on tests or lying to teachers. What they found is that correlation between different sorts of honest behavior or deceptive behavior was remarkably low, leading them to infer that the variation in behavior was better explained by variation in properties of the immediate context than by some sort of personality trait. Mischel's 1968 review of the literature on a variety of types of behavior found the same sort of pattern across the board.

Stanley Milgram's studies on obedience are an infamous part of this research tradition (1963). Milgram conducted studies putatively about learning, but which were actually about obedience to authority. Subjects were given the role of teacher in these studies, while confederates of the experimenters played the roles of learner and study administrator. The teacher's jobs were to ask questions and to administer electric shocks in response to incorrect answers (no people actually received shocks). The shocks ascended in severity in 15-volt increments. Some levels of shock were very clearly labeled with fairly dire warnings. When subjects hesitated in administering shocks, the administrator-confederate politely recited a list of instructions to continue. Milgram found that non-coercive features of experimental situations led ordinary people to administer what they thought were lethal levels of electrical shocks to other ordinary people. More precisely, about two thirds of subjects administered shocks all the way to the final level, and many of the other subjects administered shocks up to very high levels.

Overall, the situationist suggestion is that the variation in behavior exhibited by an individual should be accounted for in a way that gives a substantial role to variation in context. Various kinds of contextual property make a difference to our behavior, but a particularly important one is the presence or absence of other people and the particular relationships in which an agent stands to them. This is aptly demonstrated by the many iterations of Milgram's study.

The mere presence of others in roles of apparent authority, for one example, and the relative proximity of our 'students'/victims, for another, make a significant difference to how we behave. Crucially, these effects do not necessarily take place in a way that is mediated by conscious, deliberate thought. Social influence on action is largely unconscious. Although the details of just how such influence takes place may still need to be worked out, that our behavior is subject to surprising influence by others is not in question (for discussion see Doris 2002, Sneddon 2011).

From the standpoint of worries about autonomy, the lesson of these bodies of research might be taken to be this: much as we might value running our own lives, we are at least co-ruling creatures. If autonomy requires acting from our own values, and if values can be one's own only under conditions requiring their independence from social influence, then the results from social psychology about social influences on values compromise autonomy. If to rule oneself requires acting from one's own will, and if doing so requires independence from social influence of others, then the results of the situationist psychology compromise such rule over ourselves.

That such worries are extant is not mere speculation on my part. Examples of such intuitions in action can be found in political philosophy in general and in feminist philosophy in particular. For instance, Robert Paul Wolff explicitly deploys a version of an account of autonomy much like that involved in the intuitions indicated here: "The autonomous man, insofar as he is autonomous, is not subject to the will of another" (Wolff 1998: 14; see the section running 12–19). Much of the famous liberal-communitarian debate can be seen as a worry that liberalism presupposes an impossible ideal of individual self-rule (e. g., Sandel 1982; for summary and reflective criticism, see Sec. 2 of Bell 2016). The tradition of feminist treatment of autonomy can be roughly broken into two periods. The first was marked by suspicion of the possibility and, especially, value of self-rule given a certain understanding of it as antithetical to social relations particularly germane to the lives of women (see Friedman 2003, 36–8). The second features an explicit attempt to understand autonomy in a different, 'relational' way (e. g., Benjamin 1988, Benson 1991, MacKenzie & Stoljar 2000, Friedman 2003, Killmister 2013). In other words, the first thought in this tradition was the intuition interpreting self-rule in terms of independence from social influences that I have pointed to here. Getting over this intuition took explicit labor in the feminist tradition. In her summary of this body of work, Natalie Stoljar notes that there remains a lack of consensus about the nature of autonomy, and that the deep issue appears to be the fit of explicit accounts of autonomy with intuitions about ". . . the notion of agency that is one's own." (Stoljar 2015, Sec. 9; emphasis in orig-

inal). I concur with Stoljar about the lingering importance of attending to intuitions when addressing the nature and significance of self-rule.

The intuitive challenges to the possibility and naturalness of autonomy leave unspecified just what self-rule is. The most basic way to understand these challenges is as presuming that autonomy requires independence from others. This construal is understandable given etymology: 'self' rule is here faced by our ties to 'other' people. And, certainly, theorists of autonomy have given a great deal of attention to heteronomy, which is rule by others (Taylor 2009 is of particular note). However, once we delve beneath the surface, we can see that things are not as these challenges presume. Self-rule just does not require independence from others. With this presumption goes a great deal of the force of the intuitive challenges.

Most autonomy theorists favor what I shall call a 'structural' view according to which self-rule is held to consist in the possession, and perhaps exercise, of some psychological capacity or capacities. One reason to adopt this sort of view of autonomy is the thought that what must be explained by such a theory is just what it is for a choice (or action, or desire, or plan, and so on, but I shall focus on choices) to really be one's own, or to really come from myself, rather than coming from other people. Self-rule is, in this view, rule *by* the self. When choices issue from the right sort of psychological make-up, they are autonomous. When they don't have the right sort of psychological origin, then they are not ruled by oneself.

The most common sort of structural view is a so-called 'hierarchical' one (Frankfurt 1971, Dworkin 1988, Friedman 2003, Taylor 2005). This sort of view mobilizes the distinction between first order and higher order thoughts. A desire to visit the Dieu du Ciel brewpub in Montreal is a first order desire: it is about the world. A desire to desire to visit Dieu du Ciel is a higher order desire. Specifically, it is a second order desire. Higher order thoughts are thoughts about thoughts. Hierarchical theories of autonomy, one way or another, hold that first order thoughts are autonomous when they are endorsed by higher order thoughts. This sort of psychological structure is thought to ensure that the first order thought in question really stems from oneself rather than from some other source.

Suppose that some variety of this sort of account of autonomy is correct. For the emerging details of the extent of the social aspects of our minds to undermine self-rule, they would have to imply that the psychological structure posited as constitutive of the capacity for autonomy is necessarily ruled out by our social ties. There is no general reason to think that this is the case, and no specific consideration of which I know that implies this either. The mere fact that other people enter our psychologies in deep and surprising ways implies nothing about

our abilities to form and reason about higher order thoughts and the corresponding first order ones.

I am sympathetic to hierarchical theories of autonomy, but I favor a two-pronged account of self-rule (Sneddon 2013) Besides rule *by* the self, I think that autonomy includes both the capacity for and the exercise of rule *over* the self. Very roughly, this involves learning about ourselves, especially our motivations, thinking both about which motivations seem relatively more desirable and about the standards by which to assess the desirability of kinds of lives, and making autonomous choices in the light of such thought about what sorts of people to be. Correspondingly, our links to others would undermine rule over ourselves only if such links implied that we could not perform these sorts of thoughts. Although our relations to others can pose problems to thought, the mere fact that we are in these relations does not automatically undermine our capacities for learning about and shaping ourselves.[1]

This second prong of my favored account of autonomy introduces an important addition to purely structural accounts. Although we might well have the capacity to be autonomous persons in this sense, we are not really autonomous unless we have exercised this capacity. This means that, besides positing a collection of psychological capacities as necessary for rule over ourselves, my view also has a historical requirement. And, more generally, other theorists also supplement structural conditions of autonomy with historical ones in order to address what has become known as the "manipulation problem" (Christman 1991, Taylor 2005). Here is the manipulation problem: suppose that psychological structure X is offered as constitutive of autonomy. The worry is that X could be put in place in a manner which undermines autonomy—for instance, by a neuroscientist who tampers with one's brain without one's knowledge. In such a case, the putative psychological basis of self-rule would be present but, so the objectors hold, one would not be autonomous. The way in which X came to be undermines one's self-rule. This implies that X is not sufficient for autonomy after all. Since this problem can supposedly be raised for any psychological structure, what is needed is to supplement the structural condition for self-rule with historical conditions. These conditions specify that the psychological conditions come about in the right way(s), whatever this might be.

I am not convinced that the manipulation problem is really that much of a problem (see Sneddon 2013, 36–41) but let's grant it some force for the sake of argument. This line of thought might be pressed into service in developing the

1 Much feminist work on autonomy has been concerned with distinguishing benign from malignant sources of social influence. See, e.g., Benson 1991 & Killmister 2013.

intuitive challenges. The worry now is not that social aspects of our minds undermine autonomy by interfering with the presence of psychological capacities thought to be important to self-rule. Rather, the problem would be that our social relations imply that we acquire these capacities in ways that themselves undermine self-rule. Although we can possess the capacities for, for instance, forming and reflecting upon higher order desires, we acquire such capacities in webs of psychological dependence on other people. These connections imply that we should not think of ourselves as self-ruling even when we act from first order desires endorsed by reflective second order ones. The influence of others in the acquisition of these capacities means that their exercise fails to make the ensuing desires and actions our own. In such a complicated social world we cannot be self-ruling regardless of what our minds are like.

This is unpersuasive, and we can see why by considering more specific examples of social influence on the development of our capacities for thought. Consider educational processes. Some will indeed inhibit our abilities to think critically about choices, and, on their face, these would undermine autonomy. They do this, however, by undermining the structural basis of autonomy itself. Other educational processes are favorable to the possession and use of these capacities, which makes these supportive of self-rule. The mere fact that others help us to acquire these capacities does not rob them of their power to make us capable of running our lives.

This lesson generalizes: a general skepticism about self-rule on the basis of connections to others is unduly pessimistic. Some of our relations to others will indeed interfere with self-rule, but other social relations will enhance it. They will do this making it more the case, e.g., that our choices really come from us, or that we learn about, reflect upon, and shape our own motivations. The particular details are what matters in these cases, not the mere fact of our sociality, nor the more notable fact of the surprising scope and depth of the influence of others on our minds.

3 Extended Minds, Extinguished Selves?

Recent discussion of the metaphysics of the mind gives worries about the possibility of autonomy in the face of our sociality a novel and particular guise. The worry about sociality inherently undermining self-rule arguably turns on a deep presumption of what I shall call ‘embodied individualism’ about minds and selves. The assumption seems to be that the processes which realize autonomy must be one’s own not just psychologically but physically. That is, they must be found within the physical bounds of individual agents, such that other people

can play no legitimate role in them. They must be bodily bounded, so the assumption seems to be.

If this is what is at stake, then the extended mind hypothesis (EMH) is a source of threat to self-rule. To recapitulate the issues in a sentence: so-called individualists defend the idea that the mind must be, strictly speaking, bodily bounded, and so-called externalists deny this (see, e. g., (Clark and Chalmers 1998, Wilson 2004, Adams & Aizawa 2008, Clark 2008, Rupert 2009; Sneddon 2008, 2011). There is much work to be done to determine whether our minds really are extended, but the conceptual issues are clear: there is no *a priori* reason to think that extended cognition is impossible. Moreover, there is good reason to take extended cognitive hypotheses seriously. Brains are wonderful but not magical. In principle other things can do what brains do. Moreover, brains are evolved organs, and evolution famously makes use of whatever materials are at hand, so to speak. If resources physically external to human agents are psychologically useful, then they are within reach for the evolution of the human mind. Other people are clearly useful in lots of ways. This means that, in principle, cognition might be spread among agents rather than located within them. And this is the rub: one powerful and intuitive notion of a self is of a cognitive unity that makes one a particular person distinct from other persons. What EMH might be thought to do is to extinguish a psychologically meaningful notion of self along with the body-bounded nature of cognition. If selves are threatened by 'wide' views of the mind, so is self-rule.[2]

Here are some details about the worry. Consider Otto from Clark and Chalmers' externalist touchstone "The Extended Mind" (1998). Otto has Alzheimer's. Instead of relying on his brain for storing memories, Otto uses a notebook. He records new information there and looks up things that he has already learned. The functional role of the notebook is purported to be much the same as that of neural memory storage in healthy people.[3] On the basis of such functional parity, Clark and Chalmers claim that we should see Otto's memories, and hence beliefs, as at least partly stored in the environment beyond his body. For example, Clark and Chalmers argue that Otto has a belief about the location of the Museum of Modern Art in virtue of the information in his notebook and the way that he uses it for action (1998: 12–3). The implication is that when we attribute some belief to Otto, the location to which we are referring should not be thought to be

2 The worry about autonomy on the grounds of worries about the non-existence of the self is not peculiar to the context of debate over EMH; it gets a specific form here. It has found another form in feminist political philosophy: see Friedman 2003, 30–6 for discussion.

3 The Otto case has been the topic of much discussion. For notable dissenters see Rupert 2009, and, especially, Adams & Aizawa 2008.

Otto-the-physical-organism, but Otto-the-organism-plus-parts-of-Otto's-environment.

As presented, the Otto case involves neither social cognition nor social resources for wide cognition. Suppose then that we supplement Otto with such resources, in the spirit of the findings canvassed earlier. Imagine that Otto relies on other people for storage and retrieval of information. For some information and actions, other people play much the same functional role as his notebook (and, for others, their own healthy brains). In such a case, it looks like we should say that some of Otto's thoughts are partly stored in other people. When we attribute these beliefs to him, we aren't really referring to the organism Otto alone; we are really ascribing these thoughts to Otto-the-organism-plus-other-organisms.[4]

If this is what the combination of our sociality and the extended mind hypothesis implies, then the intuitive threat to autonomy should be clear. Autonomy is self-rule. Whatever else this involves, it presupposes some robust notion of a self that controls its life (perhaps as sketched in the previous section). For this idea to make sense, we must be able to distinguish one self from other selves and other sources of control over a life. Whatever sources of threat there are to autonomy, heteronomy is chief among them. But social versions of the extended mind hypothesis seem to blur the metaphysical distinctions between selves by loosening the functional distinctions between my mind and others' minds. If the distinction between self and other goes, with it goes the distinction between self-rule and other-rule.

So much for the apparent threat. How seriously should we take the idea that the combination of our psychological sociality and the extended mind hypothesis undermines autonomy by undermining the distinction between selves? There is much that might be said here. I shall argue that the inference from socially dependent psychological phenomena to socially extended selves is unduly hasty even for those broadly sympathetic to the extended mind hypothesis. The reasons are that there is more than one wide form that psychological hypotheses can take, and that these forms are not equally challenging to the robust psychological division between selves.

The debate over EMH concerns (at least in part) just what to make of the undeniable and multifaceted contextual dependence of much cognition. Otto's notebook is a hypothetical example of such dependence, and one to which I shall return for expository purposes, but it is worth keeping in mind the empiri-

4 Clark and Chalmers present brainstorming, interpersonal relationships marked by particularly deep dependency, and the general use of language as social manifestations of active externalism (1998: 11–2, 18–9). They briefly consider and welcome the idea that selves might be extended beyond bodies into the world without addressing the implications of social extension.

cal research pointed to earlier for real world examples. Individualists contend that the best explanation of systematic relationships between organisms and cognitive resources located outside of the organism's body will be one which locates cognition proper within the physical bounds of the organism and treats environmental resources as constitutively distinct sources of input. Such an explanation takes the individual organism as the locus of attribution of psychological states.[5] Externalists deny that the cognitive resource in question must only be such a source of input. However, this denial can come in diverse forms.

The crucial issue concerns what I shall call, for lack of a better term, the 'starting point' of psychological hypotheses. This issue is about relations between ideas, but it is useful to consider it by focusing on the psychological history of hypotheses. Rarely, if ever, are hypotheses *sui generis*. Instead they are conditioned by a background of ideas. These ideas are of various sorts: some are widely shared, some idiosyncratic, some long-standing, some new (and so on, I suspect). They also play various roles in thought. Some will be assumed, barely more than inchoate, whereas others are explicitly thought of in precise detail. Likewise, hypotheses themselves stand in various relations to such conceptual backgrounds. Some will be explicitly deduced from prior ideas whereas others will be impressionistically suggested. Imagination and luck are surely as important to science as dully grinding rigour.[6] For present purposes, the point to emphasize is that both general and particular parts of this particular background provide the starting points for hypotheses. They do this, again, in various ways, but two related ones are by making certain features of situations salient, thereby relegating others to the background, and by leading researchers to ask particular questions rather than others. With an eye on one thing rather than another, or with this question but not that one in mind, explanations can be devised, tested, improved or rejected, and so on. For instance, we should expect hard-nosed atheist materialists and believers in supernatural phenomena such as gods and souls to ask different questions and to suggest different explanations when faced with the same unusual occurrence (e. g., an earthquake, or a mass murder). There are various ways in which we might speak of these differences (the people have different outlooks, assumptions, conceptual backgrounds; they take different things for granted; they are conditioned by different concepts; they inhabit different worldviews; and more). I mean to catch all of this in the shorthand of 'starting points'.

5 If our ontology follows our best explanations, then the ontological lesson of successfully defended individualism would be that minds are individualistically realized. Likewise, the ontological lesson of externalism would be that at least some mental processes are widely realized.
6 J.D. Trout makes much of contingency for the progress of science (2016).

An important but only half-digested lesson of decades of debate over various versions of psychological externalism is that we can frame psychological hypotheses from either an individualistic or an externalist starting point even if we are defending and developing the extended mind hypothesis. Much of the literature on externalism exhibits what I have elsewhere called 'shallow' externalism (Sneddon 2008, 2011): individualistic hypotheses are taken as a starting point and reframed in an externalist fashion. This requires that psychological states be attributed to individuals as a starting point; these states are then described and explained as spread between the individual and worldly resources on the basis of systematic relations between the individual and these resources. We can contrast shallow externalist hypotheses with 'deep' ones. Deep externalist hypotheses start with the systematic relations between the individual and the relevant worldly resources and ascribe psychological states to individual agents as required for participation in this system. I describe hypotheses with this form as 'deep' with regard to their externalism because they put the wide phenomenon first. The description and explanation of the workings of individual minds follows from this starting point. The possibility of wide psychological systems is put at the foundation of this sort of hypothesis. By contrast, shallow externalist hypotheses are shallow because they have individually ascribed cognitive states at their foundation; the wide phenomena come afterwards and less fundamentally.

Let's scrutinize what might be said about Otto in the light of these distinctions. Otto stands in a systematic relationship with his notebook. What is the best explanation of the nature of the cognitive processes by which Otto uses the notebook? In particular, should we think that Otto has a belief about the address of the Museum of Modern Art in virtue of his use of his notebook? Here are three accounts:

A) Narrow Hypothesis: Otto does not have a belief about the Museum's address prior to his use of his notebook, nor one that includes the notebook as a constituent. After he consults the notebook he has a bodily-bounded belief about the museum.
B) $\text{Wide}_{\text{Shallow}}$ Hypothesis: Otto has a widely realized belief about the Museum's address. The belief is constituted by both Otto and his notebook, so it is not bodily-bounded.
C) $\text{Wide}_{\text{Deep}}$ Hypothesis: Otto and his notebook constitute a system. Otto has (at least) bodily bounded beliefs about the location (e. g., in virtue of perception) and content (e. g., in virtue of beliefs about the notebook itself, some of which are second order beliefs about, e. g., the Museum) of his notebook that give him access to widely realized information about the Museum's address via his systematic use of and reliance on his notebook. Otto does not have any first order beliefs about the Museum's address, either narrow or wide.

Hypothesis (B), the shallow externalist option, is in essence a reinterpretation of (A), the narrow option. They both start with Otto as a locus of attribution of psy-

chological states which are cast as respectively narrow and wide. Imagine researchers whose starting point for the explanation of behavior is the attribution of cognitive capacities to individuals. They observe Otto and ask such questions as 'What is Otto (the individual) doing?' 'What capacities are needed for this?' The individualistic researchers answer these questions in one way, the externalist ones in another, but they are asking essentially the same questions. By contrast, (C) starts with the systematic relationship between Otto and the notebook and attributes psychological states to Otto as needed to make use of the notebook. Imagine researchers here asking such questions as 'Are there systems here? If so, how is Otto engaged in these systems? What capacities are needed for this engagement?' This is a deeply externalist option because its starting point is not an individual but a system that includes an individual as a part but that is not limited to that individual.

In this case, there is good reason to prefer (B) to (C) [and to (A), but the important contrast for my purposes is between the wide hypotheses]. The reasons are those that Clark and Chalmers offer (1998: 12–6). However, in principle there can be cases better explained by deeply wide hypotheses than by shallow ones. To see this, here is another toy case. Marilyn and Susan are chatting and strolling around their university campus when they encounter some robots that they have never seen before. Some of the robots are red, some are green. There are many more green robots than red ones. They are all digging in an open field. They change location occasionally, stopping then restarting their digging. After watching for a while, Marilyn and Susan realize that the red robots are always the first to move and that the green ones follow shortly afterwards, making comparable adjustments in their location.

Marilyn and Susan start to discuss hypotheses about how the robots work. All of the robots clearly have the same moving and digging equipment, but they seem to differ in other ways. Marilyn suggests that they are digging for something specific. She thinks that the red ones must be equipped with two sorts of devices, one for detecting the intended target and the other for transmitting location information to the green ones. The green ones, she argues, have receivers: once they receive location information from the red robots, they adjust their location and recommence digging.

Susan agrees that the red robots must have a device for detecting whatever they are digging for, but she disagrees about the hypothesis about the transmitters and receivers. She doesn't see why the red robots must have special devices for sending out messages. Instead, she thinks that the green robots have a way of directly utilizing the information that the detection devices realize. It is as if the green robots read the minds of the red ones; the red ones don't have to send out messages for this to happen. Susan thinks that this might be done indirectly, via

the formation of representations of the substance's location by inferences that the green robots make based on the red robot's behavior. However, she favors a hypothesis that involves more direct mind-reading: she imagines that the green robots have something akin to a psychological x-ray that allows them to access the representations inside the red robots without the red robots having a mechanism for sending this information to them.

As Marilyn and Susan chat about the robots, they are overheard by Gabrielle. Gabrielle interrupts the conversation and introduces herself: she is one of the designers of the robots, so she can settle Marilyn and Susan's dispute. As it happens, neither of them is quite right. The red robots do have a detection device for the targeted substance. This device is very expensive to make, so the engineers wanted to minimize the number that they had to construct. The red robots do not have transmitters, but the green robots don't read the detection device directly either. In fact, the green robots do not have representations of the location of the target substance at all. Instead, the green robots have very cheap mechanisms for mimicking the movements of the red robots, along with some software that keeps them a practical distance apart. At first the engineers expected that this arrangement might work and that it would save them a little money, but that it would be second-best to having a full squad of the red 'detector' robots. But once they tried out the robots, they were pleasantly surprised by how successful this setup was. Consequently, they added many more of the cheap green 'mimic-bots', resulting in the coordinated behavior observed by Marilyn and Susan.

Marilyn and Susan share the same starting point with regard to the formulation of psychological hypotheses. They both focus on individuals and the attribution of capacities to them. Marilyn's hypothesis is an individualistic one. She posits representations of the location of the substance within the red robots due to their detection devices and within the green robots due to the transmission of this information from the red robots to the green ones. Susan's hypothesis is a shallowly wide one. Besides the red robots' narrowly realized representations of the location of the substance due to the detection device, she supposes that the green robots have widely realized representations of this substance. The wide realization is constituted by the information inside the red robots and the 'mind' reading capacities of the green robots that give them access to the location information.

The truth as revealed by Gabrielle is different, so neither the narrow hypothesis nor the shallowly wide one is correct. Instead, a deeply wide hypothesis is needed to get at the truth. Put otherwise, a starting point that oriented researchers toward inter-individual systems rather than to the individuals independently of systematic interaction raises the right sort of questions for identifying the actual mechanisms responsible for the robot behavior. The green robots' behavior is pro-

duced due to the information about the location of the substance that the red robots have, but the green robots do not re-produce or directly access this information themselves. Instead, they participate in a wide system that allows this information to inform their behavior. That this is reasonably seen as widely systematic information processing is bolstered by the fact that the capacities and activities of the robots are responsible for the proliferation of particular kinds of capacities through the robot population. Specifically, the success of the green mimic-bots at finding the targeted substance is the reason that the engineers made so many of them. All that the green robots need in order to participate in this system and to have their behavior directed by information about the location of the target substance is the ability to mimic the movements of the red robots. This ability is individualistically located. Functionally, the green mimic-robots participate in a wide information-processing system in virtue of physically bounded cognitive capacities. Likewise, the cognitive capacities of the red robots are also individualistically located. The wide information-processing system is instantiated by the green and red robots entirely in virtue of bodily-bounded capacities.

The robot case is offered as a model for a certain sort of psychological hypothesis that can be devised to explain real-world behavior.[7] Suspicions about extended cognition in principle need not only be developed from otherwise individualistic starting points; the starting point for wide hypothesis formation can itself be wide. With regard to self-rule in particular, the important thing to see about this toy case is that it combines wide cognitive systems with a robust sense of individually located cognition. Because of this it offers a way of pursuing the ideas that motivate EMH while preserving a functionally robust notion of selves. Wherever deeply wide hypotheses fit us, we find the combination of externalism and a meaningful model of a self.

Two complications are worth noting in passing. First, real-world bio-minds such as our own do not have minded engineers. We are the products of the mindless design processes of natural selection. Hence when framing and evaluating real psychological hypotheses, our reasons for thinking that we find systematic mind-world integration will differ from those found in this toy case. This is a point of detail, not a deep disanalogy. Particularly important will be phenomena characterized by individual-resource coupling, especially between individuals, that is responsible for the persistence and even proliferation of a behavior (just as in the robot case: the engineers reproduce and spread the green robots through the robot population due to the behavioral success born of their mimicry

7 And, on the assumption that ontology follows explanation, as a model for understanding the constitution of real-world minds.

of the red robots).[8] Second, and more importantly, there is no reason for human minds not to be explicable in terms of both shallow and deep externalist hypotheses. The plurality of cognitive mechanisms that we exhibit invites a deeply diverse array of explanation, some wide, some narrow, some deeply wide and some only narrowly so.

Let's return to humans. The lesson is this: when we are impressed by the environmental dependence of human cognition, including its social aspects, we have (at least) two choices. We can attempt to explain this through reinterpretation of individualistic understanding of our minds, or we can prioritize the wide phenomena. The first route threatens autonomy by spreading minds between people and thereby extinguishing boundaries between self and others. The second route preserves a psychologically rich notion of individual agents—of selves—while allowing for extended cognitive processes. This in turn defuses the supposed threat to autonomy posed by the extended mind hypothesis, as this threat turned on a worry that EMH made problematic the distinction between oneself and other minds. Whether deeply wide hypotheses are indeed true of us is an empirical matter. Nevertheless, the mere conceptual possibility of extended cognition should not be thought to be a particular threat to human autonomy.

The ruminations about narrow and wide hypotheses have been aimed at a worry about EMH and the 'self' component of self-rule. However, as wide and narrow hypotheses about particular sorts of cognition are formed, tested, refined, retested, and so on, there is reason to expect lessons about the 'rule' component as well.[9] Generally, once people start to frame new sorts of hypotheses,

8 See Sneddon (2011: 17–23), for discussion of (hypothetical) avian cognition and natural selection as a model for wide systematic human-human cognitive processing.

9 Consider a very ordinary example of the forms which extended cognition might take. One of the most obvious ways in which we gain access to other people's minds is by talking to them. Conversation facilitates access to and critical thought about our own thoughts, such as the choices we make. People famously seek out others to talk through tricky decisions in their own lives. Why? There are various things that can be happening here. One is that conversation makes our ideas and options public for the attention of all, including ourselves. A second is that communication requires the relatively clear formulation of ideas, often going beyond what we do for ourselves alone. A third is that we gain the perspective of another person, or of multiple people, on ourselves and our options when we talk about ourselves. A fourth is that reasoning actually takes place in the back-and-forth manipulation of ideas that constitutes a conversation.

The kinds of processes that such conversations enact might well realize one's autonomy. For example, in talking we might give our first order motivations a symbolic formulation that we had not given them before. Working over these public symbols can lead to new higher order motivations, thereby shaping both what we desire and what we eventually do. Since conversation is interpersonal, this way of realizing our autonomy requires other people. In their absence, our possession and exercise of the capacities constitutive of self-rule would degrade. Still,

they tend to have new sorts of thought. This means that they notice things that they hadn't seen before and think of what had seemed familiar in novel ways. Rather little work in psychology has been done in an explicitly externalistic vein (still, it's more common than when I started thinking about these issues). Here is a speculative wager: the mechanisms for social conformity that Milgram infamously studied and about which social psychologists have had much to say will turn out to have surprising effects for our autonomy. I mean this in a positive sense, not (just) a deleterious one.

Second, and relatedly, once we start to think about wide cognitive systems, we gain the possibility of seeing and exploiting novel wide cognitive opportunities. Even further, we get the possibility of designing them for ourselves. To the extent that we can devise novel ways of, and even systems for, relying on each other, we can, in principle, facilitate self-rule widely. After all, there is no reason to think that our capacities for self-rule are only either fixed or subject to threat. The perspective of the extended mind hypothesis surely offers promise at least as much as it does despair.

References

Adams, Frederick and Aizawa Kenneth. 2008. *The Bounds of Cognition*. Oxford: Blackwell.

Bell, Daniel. 2016. "Communitarianism". *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), Edward N. Zalta (ed.). <https://plato.stanford.edu/archives/sum2016/entries/communitarianism/>.

Benjamin, Jessica. 1988. *The Bonds of Love: Psychoanalysis, Feminism, and the Problem of Domination*. New York: Pantheon Books.

Benson, Paul. 1991. "Autonomy and Oppressive Socialization". *Social Theory and Practice*. 17. 385–408.

Christman, John. 1991. "Autonomy and Personal History". *Canadian Journal of Philosophy*. 21. 1. 1–24.

Clark, Andy. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Clark, Andy and Chalmers David. 1998. "The Extended Mind". *Analysis*, 58. 10–23.

Craig, Joseph and Haidt Jonathan. 2007. "The Moral Mind: How 5 Sets of Innate Moral Intuitions Guide the Development of Many Culture-specific Virtues, and Perhaps Even

the first order and higher order motivations are one's own, in the sense of (apparently) being directly and legitimately attributable to us as individual agents. Moreover, one participates in the public reasoning process not just as an equal partner but as a participant with a special interest in the topic, which is one's own motivations and actions. There is every reason to see such processes as both interpersonal and partly constitutive of personal autonomy.

Modules". In Peter Carruthers, Stephen Laurence and Stephen Stich (eds.) *The innate mind*. Vol. 3.

Doris, John. 2002. *Lack of character*. Cambridge: Cambridge University Press.

Dworkin, Gerald. 1988. *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.

Frankfurt, Harry. 1971. "Freedom of the will and the concept of a person". *Journal of Philosophy* 68. 5–20.

Friedman, Marilyn. 2003. *Autonomy, Gender, Politics*. Oxford: Oxford University Press.

Haidt, Jonathan, Koller Silvia Helena and Dias Maria. 1993. "Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?". *Journal of Personality and Social Psychology*. Vol. 65, No. 4. 613–28.

Haidt, Jonathan and Bjorklund Fredrik. 2008. "Social Intuitionists Answer Six Questions About Moral Psychology". In Walter Sinnott-Armstrong (ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. Cambridge, (MA): MIT University Press.

Hartshorne, Hugh and May Mark A. 1928. *Studies in the Nature of Character I: Studies in Deceit*. New York: MacMillan.

Iyer, Ravi, Koleva Spassena, Graham Jesse, Ditto Peter and Haidt Jonathan. 2012. "Understanding Libertarian Morality: The Psychological Dispositions of Self-identified Libertarians". *PLoS ONE* 7(8): e42366. https://doi.org/10.1371/journal.pone.0042366.t007

Killmister, Suzy. 2013. "Autonomy and the Problem of Socialization". *Social Theory and Practice* 39 (1). 95–119.

Milgram, Stanley. 1963. "Behavioral Study of Obedience". *Journal of Abnormal and Social Psychology. 67*. 371–378.

Mischel, Walter. 1968. *Personality and Assessment*. New York: John Wiley and Sons.

Rozin, Paul, Lowery Laura, Imadam Sumio and Haidt Jonathan. 1999. "The CAD Triad Hypothesis: A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity)". *Journal of Personality and Social Psychology*. Vol. 76 (1999). 574–86.

Rupert, Robert D. 2009. *Cognitive Systems and he Extended Mind*. Oxford: Oxford University Press.

Sandel, Michael. 1982. *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press.

Shweder, Richard, Much Nancy, Mahapatra Manamohan and Park Lawrence. 1997. "The 'Big Three' of Morality (Autonomy, Community, Divinity) and the 'Big Three' Explanations of Suffering," In Allan Brandt and Paul Rozin, (eds.), *Morality and Health*. New York: Routledge.

Sneddon, Andrew. 2008. "The Depths and Shallows of Psychological Externalism." *Philosophical Studies*. Vol. 138, No. 3 (April). 193–208.

Sneddon, Andrew. 2011. *Like-Minded: Externalism and Moral Psychology*. Cambridge, (MA): MIT Press.

Sneddon, Andrew. 2013. *Autonomy*. London: Bloomsbury.

Stoljar, Natalie. 2015. "Feminist Perspectives on Autonomy". *The Stanford Encyclopedia of Philosophy (Fall 2015 Edition)*, Edward N. Zalta (ed.). <http://plato.stanford.edu/archives/fall2015/entries/feminism-autonomy/>.

Taylor, James Stacey. (ed.) 2005. *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press.
Taylor, James Stacey. 2009. *Practical Autonomy and Bioethics*. New York: Routledge.
Trout, J.D. 2016. *Wondrous Truths: The Improbable Triumph of Modern Science*. Oxford: Oxford University Press.
Wilson, Robert. 2004. *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition*. Cambridge: Cambridge University Press.
Wolff, Robert Paul. 1998. *In Defense of Anarchism*. Berkeley: University of California Press.

José V. Hernández-Conde

Articulating Context-Dependence: *Ad Hoc* Cognition in the Prototype Theory of Concepts

Abstract: Recently, Casasanto and Lupyan (2015) have proposed an appealing and daring thesis: there are no context-independent concepts—that is, all concepts are *ad hoc* concepts. They argue that the seeming stability of concepts is merely due to commonalities across their different instantiations but that, in fact, there is nothing invariant in them. In their view, concepts only exist when they are instantiated for categorizing, communicating, drawing inferences, etc., and those instantiations are produced on the fly from a set of contextual cues. However, the main weakness of Casasanto and Lupyan's framework is that it lacks a proposal for articulating it within a theory on the structure of concepts. My aim is to show that the *ad hoc* cognition framework can be characterized by means of a prototype theory of concepts developed in terms of a conceptual similarity space.

1 Concepts and Context-Dependence

Concepts play a key role in cognitive processes such as categorization, inference, learning, memory, decision making, problem solving, etc., being commonly identified with bodies of knowledge about the members of a given category. The traditional view identifies concepts with cores of knowledge stable across individuals and time (*invariantism*) (Machery 2009), which explains both the accumulation of knowledge by individuals and the ability to communicate with other subjects. Other views argue that many concepts are context-dependent construals created on the fly (*contextualism*) (Barsalou 1993; Sperber and Wilson 1995; Carston 2002; Prinz 2002; Malt 2010), which explains our adaptive behavior to changing environments.

In the contextualist approach Casasanto and Lupyan (2015) have proposed the thesis that all concepts depend on context—or, in other words, that all concepts are *ad hoc* concepts. They convincingly argue that it is necessary to abandon the idea that concepts have stable or default cores accessed by people when they instantiate those concepts. In their view, evidence of this would be the impos-

José V. Hernández-Conde, University of the Basque Country (jhercon@gmail.com)

https://doi.org/10.1515/9783110702286-008

sibility to draw a boundary between the core—that is, a set of stable and context-independent properties accessed by subjects whenever they instantiate a concept—and the periphery of a concept. Incidentally, that was precisely what Wittgenstein (1953: § 66–100) showed in his discussion of family resemblance: it is not possible to identify a common property—or set of properties—for all the things we call "game"; or, in other words, there is nothing invariant in concepts.

As a matter of fact, Casasanto and Lupyan maintain that, given that the subject's cognitive state is a part of the context, and considering that the brain is continuously changing, this would entail that concepts are inherently variable. If Casasanto and Lupyan are right, then concepts would only exist when they are instantiated for categorizing, communicating, drawing inferences, etc., and those instantiations would be produced on the fly from a set of contextual cues.

However, after asserting that *concepts are not something we have in the mind, but something we do with the mind* (Casasanto & Lupyan 2015: 246), Casasanto and Lupyan focus their work on the instantiation of concepts, leaving aside the issue of what cognitive structures can ground those instantiations. Nevertheless, in order to accept the *ad hoc* cognition framework, a characterization of the cognitive structures supporting the instantiation of concepts is demanded. The rest of this work is devoted to the issue of how Casasanto and Lupyan's approach might be articulated within a theory on the structure of concepts, paying special attention to the question of how the context-dependence of every instantiated concept may be put in place. I will try to show that *ad hoc* cognition can be conceived in terms of a prototype theory of concepts built over a geometrical similarity space.

2 Prototype Theory and *Ad Hoc* Cognition

According to the prototype theory, concepts are prototypes, that is, representations whose structure encodes information about the properties that their members tend to have. Howbeit, there are distinct ways in which the prototype theory can be articulated (Smith & Medin 1981):

(a) *Featural models:* an object *O* is classified under a concept *C* if it possesses a sufficient number of the properties associated to *C*.
(b) *Dimensional models:* an object *O* is classified under a concept *C* if it *possesses to some degree* a sufficient number of those properties.

In both cases an object *O* will be categorized or not under a particular concept *C* in function of the similarity between *O* and the prototype of *C*, which will be determined by virtue of their shared properties. If, in the case of dimensional mod-

els, the objects and the prototypes of concepts are represented in a geometrical space whose dimensions are the constitutive properties of the relevant concepts —for the considered context—, then that would be what is known as a *similarity space theory of concepts*.

2.1 Similarity space theories of concepts

A similarity space theory of concepts can be described by the next fundamental thesis (Gauker 2007): the mind is a representational hyperspace within which (a) *dimensions* represent ways in which objects can differ, (b) *points* represent objects, (c) *regions* represent concepts, and (d) *distances* are inversely proportional to similarities—between objects or concepts. In consequence, an object *O* will belong to a concept *C* if and only if its values in every dimension of that similarity space produce an *n*-tuple that lies inside the region associated with *C*.

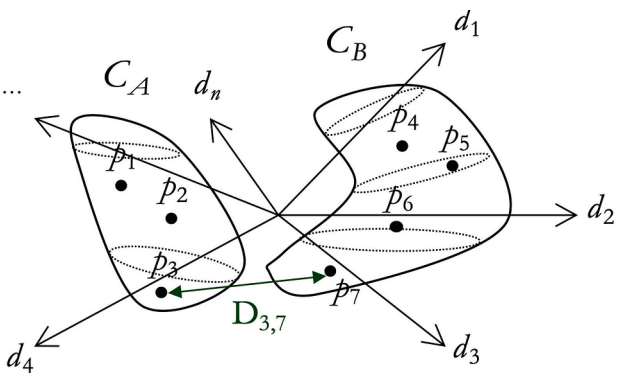

Figure 1: Illustrative example of a conceptual similarity space.

For instance, Figure 1 shows a conceptual similarity space constituted by n dimensions d_i, where the concepts A and B are represented by the regions C_A and C_B. The points p_j represent distinct objects, three of which (p_1 to p_3) are categorized under the concept A, while the other four (p_4 to p_7) under the concept B. The similarity between two objects—p_3 and p_7, for example—would be inversely proportional to the distance between them ($D_{3,7}$).

The prototypes of concepts would result from a process of maximization of similarities (or, alternatively, minimization of distances) between the evaluated objects, and the tentative prototypes. The set of final prototypes will be the one that maximizes intra-group similarity and minimizes inter-group similarity.

Thence, the prototype of a concept arises as the generalization of the properties of the objects chosen as tentative members of its associated category.

Lastly, the shape and boundaries of the conceptual regions may be the result of a Voronoi tessellation of the conceptual hyperspace, whose input were the prototypes of the set of relevant concepts (see Figure 2).

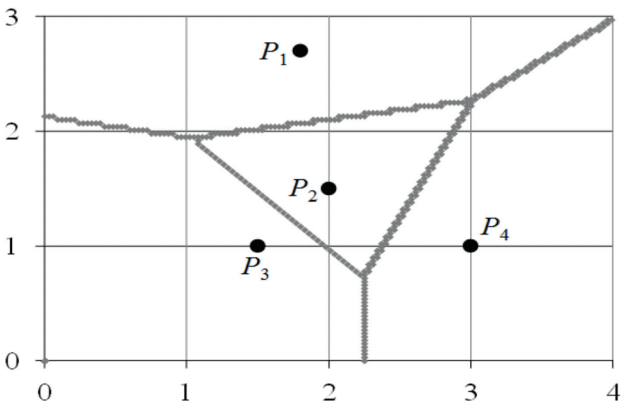

Figure 2: Boundaries of the conceptual regions resulting from the tessellation of a Euclidean conceptual hyperspace, by means of a maximization process following the principles of the prototype theory of concepts. The final prototypes are represented by four black dots with coordinates (1.5,1), (1.8,2.7), (2,1.5) and (3,1). The boundaries of the conceptual regions are represented by means of grey dotted lines.

2.2 The Distinction between Prototypes and Conceptual Regions

Advocates of similarity spaces sometimes identify concepts with prototypes or conceptual regions indistinctly, as Gärdenfors does with his definition of (natural) concept in terms of a set of conceptual regions:

Criterion C: A natural concept is represented as a set of regions in a number of domains together with an assignment of salience weights to the domains and information about how the regions in different domains are correlated. (Gärdenfors 2000: 105)

Here my point is that regions and prototypes are very different things, and that concepts must be identified with the prototypes—and not with the regions. Indeed there are significant reasons that support these statements:

— What results from the generalization of a set of tentative examples of a given category is a prototype, not a region. Conceptual regions only arise from the

evaluation of the distances between all the points of the conceptual hyperspace, and the prototypes of the relevant concepts.

– The application of conceptual regions in categorization tasks is both unnecessary and inefficient: (A) It is *unnecessary* because in order to categorize an object only the locations of the relevant prototypes are needed. (B) It is *inefficient* —both in terms of memory and/or processing—because it compels either to store the concept associated to every point, or to store all the boundaries and determine the region within which the considered object is situated.

Therefore, it is an error to attribute to the conceptual regions a persistent and strong ontological sense. (Their function is merely explicative, because it is easier to say that 'an object *O* falls within the region associated to a concept *C*', than to say that 'the distance between *O* and the prototype associated to *C* is less than the distance between *O* and the prototype of any other concept distinct to *C* (and relevant for that context)'.

In consequence, the *information stored* by our cognitive system about concepts would be the locations of their prototypes, and not their associated regions and/or boundaries.

2.3 A model for *ad hoc* cognition

Now let us see how the instantiation of a concept within a similarity space theory could take place. Remember that in this kind of theories similarity is inversely proportional to the distances between objects and/or the prototypes of concepts. Thence, the Minkowski distance between two concepts (and/or objects) *A* and *B*, located within an *n*-dimensional space, would be given by the following expression:

$$d(A,B) = \left(\sum_{i=1}^{n} w_i \left| X_i^{[A]} - X_i^{[B]} \right|^p \right)^{1/p}$$

where, $X_i^{[Y]}$ represents the value of the *i*-th dimension of the object or concept *Y*; w_i represents the weight assigned to the contribution of the *i*-th dimension; and the value of the parameter *p* determines the kind of metric (e. g., if p=1 the metric is city-block or Manhattan; if p=2 the metric is Euclidean).

The expression above applies to the standard Minkowski distance. However, those distances might be weighted differently according to various criteria. For instance, the weight could be a function of the number of examples on which a given concept is based. In such a case, the distance-of-comparison

$d_{C_k}(O, P_{C_k})$ between an object O and a concept C_k, may be expressed under a multiplicatively weighted scheme (Okabe *et al.* 1992: 119 – 134):

$$d_{C_k}(O, P_{C_k}) = u_k d(O, P_{C_k})$$

where, u_k represents the weight assigned to the distances from C_k.

In consequence, the categorization of an object O under a particular concept is the result of a cognitive process that (1) evaluates the distances from O to the prototypes of all the relevant concepts—within the considered context—, and (2) classifies O under the closest concept—and thus most similar to O.

Inasmuch as distances—and similarities—are a function of the parameters p, w_i and u_k, and given that the categorization of an object depends on which the relevant concepts are, there would be at least four context-dependent factors that can influence the instantiation of every concept in a characterization of the *ad hoc* cognition framework like this:

- the instantiated concepts (see Figure 3),
- the kind of metric—parameter p—(see Figure 4),
- the importance of dimensions—weights w_i—(see Figure 5), and
- the significance of concepts—weights u_k—(see Figure 6).

Thence, each new instantiation of a concept in a particular context would be different, given that the relevant concepts, the kind of metric, and the importance of dimensions and concepts will vary from context to context.

In consequence, a prototype theory of concepts (conceived in terms of a geometrical similarity space) can provide a successful account of Casasanto and Lupyan's main thesis, that is, that all concepts are *ad hoc* concepts—or, in other words, that the instantiation of every concept depends on the context in which such an instantiation happens.

3 Life Cycle of a Concept

Hitherto, two distinct notions of concept have been tacitly used in the previous sections. Now I will clearly distinguish those two different notions of concept, which may be identified with two distinct stages in the life cycle of a concept.

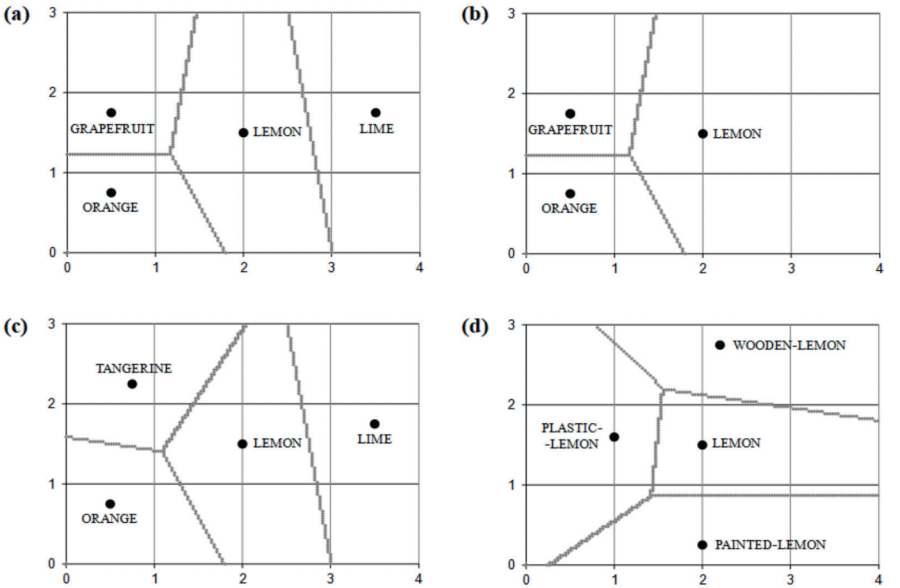

Figure 3: Example of contextual dependence of concepts due to the set of relevant instantiated concepts, for a categorization process of citruses where abscissas may be identified with *color*, and ordinates with a mixture of *texture* and *shape*. **a** Default context with prototypes of the concepts LEMON, ORANGE, GRAPEFRUIT and LIME located—respectively—in the coordinates (2,1.5), (0.5,0.75), (0.5,1.75) y (3.5,1.75). **b** Context where the concept LIME is not relevant. **c** Context where the third relevant concept was not GRAPEFRUIT, but the concept TANGERINE, located in (0.75,2.25). **d** Context where the relevant concepts were LEMON, PAINTED-LEMON, PLASTIC-LEMON and WOODEN-LEMON, the last three located in (2,0.25), (1,1.6) y (2.2,2.75)—respectively.

3.1 Concepts as storage

The first notion of concept would be associated with the information stored within our cognitive system regarding a given category. From here on I will refer to them as *stored concepts* or *concepts as storage*.

In the case of my proposal, that is, a prototype theory of concepts built over a geometric similarity space, the only information that needs to be registered by our cognitive system is the location of the prototype associated to each concept. Such locations are the only thing required to instantiate a concept within a particular context—that is to say, to determine the distances and similarities between that concept and any other object or concept. Therefore, *stored concepts*

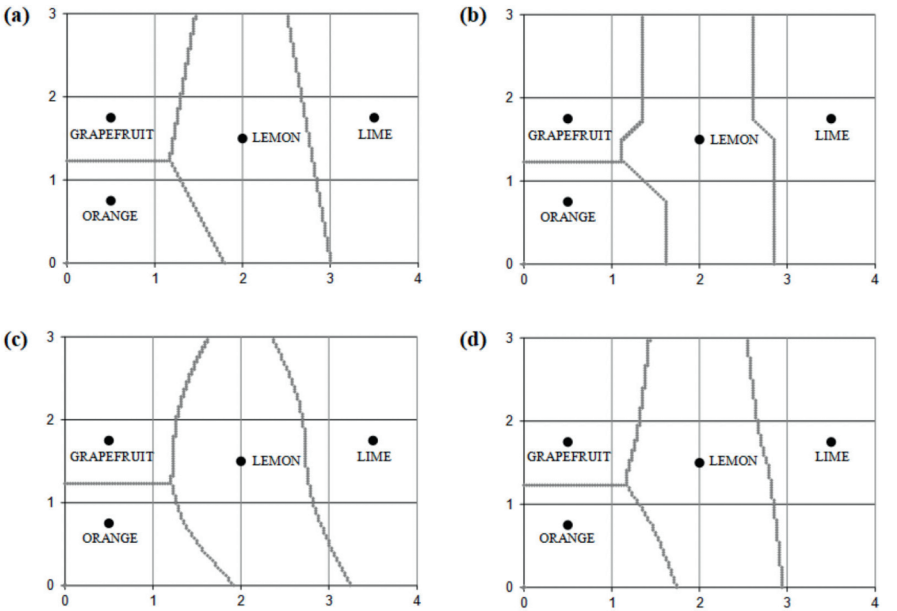

Figure 4: Example of contextual dependence of concepts due to the kind of metric, for a categorization process of citruses with the same set of relevant concepts located—respectively—in the coordinates (2,1.5), (0.5,0.75), (0.5,1.75) y (3.5,1.75). **a** Context with Euclidean metric (p=2). **b** Context with city-block/Manhattan metric (p=1). **c** Context with higher-order metric (p=3). **d** Context with optimal metric for integral dimensions (Handel and Imai 1972) (p=1.7).

are the information persistently registered by our minds about the location of their prototypes.

Additionally, *stored concepts* provide the continuity needed to accumulate new information over time about a given category (e. g., when, as a result of subsequent executions of the learning processes, new properties are added to the previously stored location of its prototype). The advantage of this is that, from a radical contextualist approach, it is possible to explain a typically invariantist ability—to wit, the accumulation of new knowledge by individuals.

However, and although the stored concept is the starting point for any instantiation of a concept—which may take place in cognitive processes such as categorization—, the stored concept cannot determine the output of those processes by itself, because additional contextual factors are involved in them. Remember that the instantiation of a concept requires the calculation of distances / similarities between the evaluated object and the prototypes of all the context-relevant concepts, and that that computation depends on the kind of metric, the importance of dimensions, and the significance of the considered concepts.

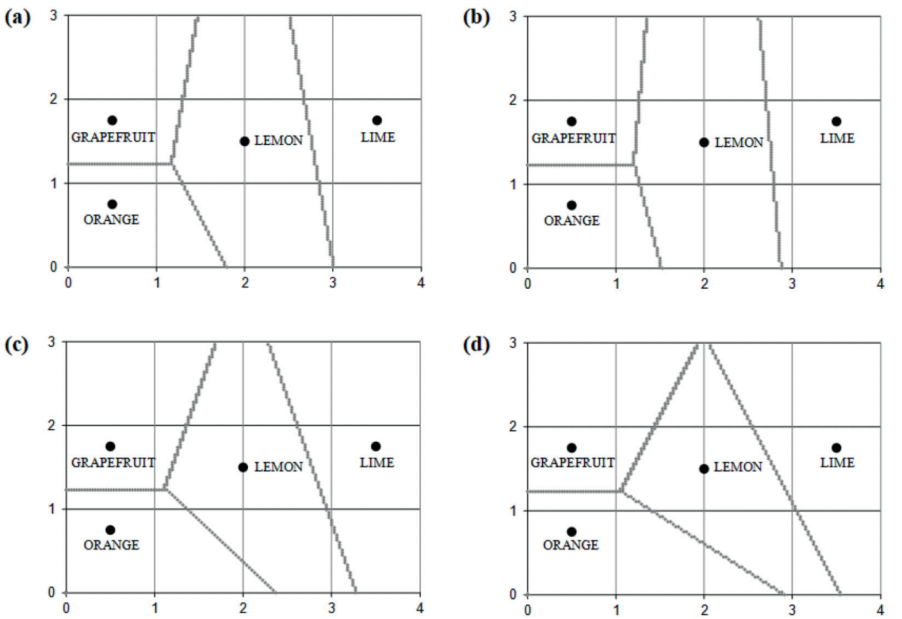

Figure 5: Example of contextual dependence of concepts due to the weight of dimensions, for a categorization process of citruses with Euclidean metric, based on *color* (abscissas) and a mixture of *texture* and *shape* (ordinates). **a** Default context with equally weighted dimensions [weights (1,1)]. **b** Context where *color* had twice the weight of *texture* and *shape* [weights (2,1)]. **c** Context where the mixture of *texture* and *shape* had twice the weight of *color* [weights (1,2)]. **d** Context where the mixture of *texture* and *shape* had thrice the weight of *color* [weights (1,3)].

3.2 Concepts as instantiation

As said above, the mere information stored about a concept does not explain how that concept is used in tasks such as categorizations, inferences, etc. The reason is that what is involved in those cognitive processes is not the stored concept, but the instantiation of that information—*instantiated concept* or *concept as instantiation*—which will depend on the context. The instantiated concepts can be identified with the *ad hoc* concepts proposed by Casasanto and Lupyan, that is, they might play the role attributed to concepts by a radical contextualist approach.

However, the idea of *instantiated concept* is more slippery than the notion of *stored concept*. This is so because stored concepts can be thought to be persistently backed by a certain structure, either informational—record system, neural

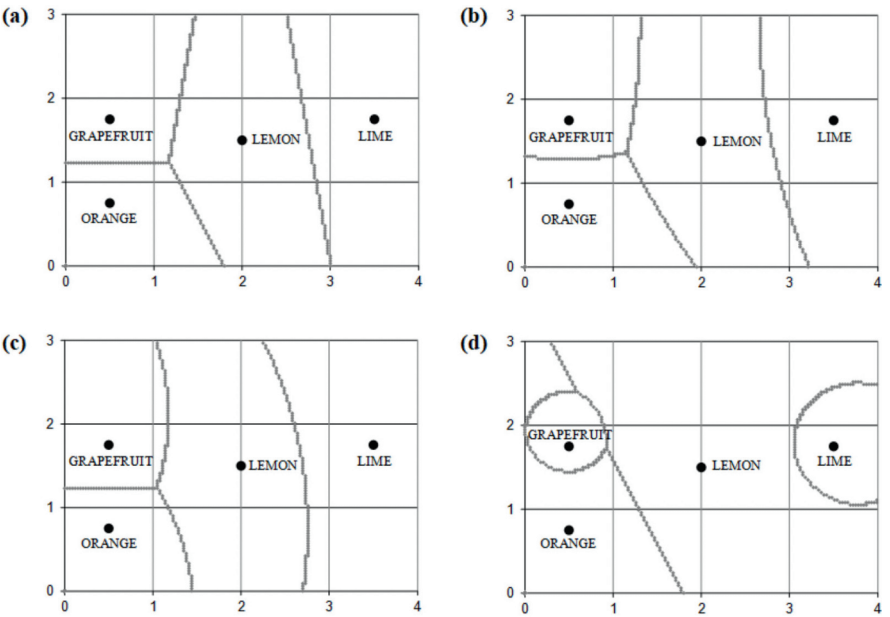

Figure 6: Example of contextual dependence of concepts due to the different significance of the relevant concepts, for a categorization process of citruses. **a** Default context with equally weighted concepts [weights (1,1,1,1)]. **b** Context with concepts weighted by their relative frequency [weights (1.1,1.2,1,1)]. **c** Context for a worker in a production line of lime nets [weights (1.3,1,1,1.5)]. **d** Context for a child who had been exposed to very few examples of grapefruits and limes [weights (2.5,2.5,1,1)].

network, etc.—or physical—potential level, electrochemical gradient, etc.—, but an identification like this is not possible for the instantiated concepts.

The reason for this is that the instantiated concept is produced on the fly depending on the subject's context, every time a concept is considered relevant for a categorization process. Therefore, *concepts as instantiations* exist only as a result of cognitive processes associated with categorizations, inferences, etc., in spite of which they are responsible for the external manifestation of those concepts (in fact, the result of those processes is the only sort of empirical evidence we have about what we call 'concepts'). Hence, the instantiated concepts are not something that exist; conversely, they are something that happen at the end of that kind of cognitive processes.

3.3 Two stages in the life cycle of a concept

Finally, my view is that *storage* and *instantiation* are not distinct notions of concept resulting from alternative theories about what concepts are, but different stages within the life cycle of a concept. In the first place, when a concept *C* is acquired our cognitive system stores certain information about it. Under a similarity space theory of concepts, that information (or *stored concept*) would be the location of the prototype associated with *C*, which is registered in a stable and persistent way until new perceptual information triggers a revision of that concept. At a later time, when *C* is used as a tentative concept for the categorization of an object under a particular concept, part of the information stored about *C* is read from memory, together with information stored about other concepts relevant in that particular context and other context-dependent factors. This last kind of cognitive processes gives rise to the so-called *instantiated concepts*, which are absolutely dependent on the actual context and that, due to it, cannot exist before all the contextual factors are determined from that context—that is to say, instantiated concepts only exist at the end of the processes of categorization, inference, etc. that instantiate them.

Obviously, this sort of life cycle of a concept is not linear but circular, because: (i) part of the information stored about concepts is used in order to instantiate them, thus the instantiated concepts depend on the stored concepts; and (ii) the categorizations of objects resulting from different instantiations of a concept will be used by subsequent executions of the acquisition processes, which leads to the modification of the existing concepts.

4 Conclusion

In this work I have tried to show that Casasanto and Lupyan's *ad hoc* cognition framework can be characterized by means of a prototype theory of concepts developed in terms of a geometrical similarity space. My proposal is compatible with Casasanto and Lupyan's thesis that there are no context-independent concepts, and—in the pages above—I have identified four possible sources of contextual dependence: relevant concepts, kind of metric, importance of dimensions, and weights of concepts.

Additionally, two different notions of concept have been distinguished, associated with two distinct stages of their life cycle: (a) *concepts as storage*, or information persistently stored by the mind about concepts, and stable between different executions of the concept-acquisition processes; and (b) *concepts as instantiations*, or the ones responsible for the external manifestation of con-

cepts—such as categorizations and inferences—, which only exist when they are instantiated in those cognitive processes.

The major advantage of this approach is that it brings together virtues both from the contextualist and invariantist approaches. First, it articulates a contextualist framework compatible with the evidence against the existence of definitions (or conceptual cores), and thus able to provide an account of our adaptive abilities to changing environments. Secondly, stored concepts are stable enough to explain how new information on them is collected.

References

Barsalou, Laurence W. 1993. "Flexibility, Structure, and Linguistic Vagary in Concepts: Manifestations of a Compositional System of Perceptual Symbols". In Alan F. Collins, Susan E. Gathercole, Martin A. Conway, and Peter E. Morris (eds.). *Theories of Memory*. Hillsdale: Lawrence Erlbaum Associates. 29–101.

Carston, Robyn. 2002. *Thoughts and Utterances*. London: Blackwell.

Casasanto, Daniel and Lupyan Gary. 2015. "All concepts are ad-hoc concepts". In Eric Margolis and Stephen Laurence (eds.), *The Conceptual Mind: New Directions in the Study of Concepts*. Cambridge (MA): MIT Press. 543–566.

Gärdenfors, Peter. 2000. *Conceptual Spaces: The Geometry of Thought*. Cambridge (MA): MIT Press.

Gauker, Christopher. 2007. A Critique of the Similarity Space Theory of Concepts. *Mind & Language* 22. 317–345.

Handel, Stephen and Imai Shiro. 1972. "The Free Classification of Analyzable and Unanalyzable Stimuli". *Perception and Psychophysics* 12. 108–112.

Machery, Edouard. 2009. *Doing Without Concepts*. Oxford: Oxford University Press.

Malt, Barbara C. 2010. "Why We Should Do Without Concepts". *Mind & Language* 25. 622–633.

Okabe, Atsuyuki, Boots Barry, Sugihara Kokichi and Chiu Sung N. 1992. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. New York: John Wiley & Sons.

Prinz, Jesse J. 2002. *Furnishing the Mind*. Cambridge (MA): MIT Press.

Smith, Edward E. and Medin Douglas L. 1981. *Categories and Concepts*. Cambridge (MA): Harvard University Press.

Sperber, Dan, and Wilson Deirdre. 1995. *Relevance: Communication and Cognition* (2nd ed.). Oxford: Blackwell.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*, G.E.M. Anscombe & R. Rhees (eds.), G.E.M. Anscombe (trans.). Oxford: Blackwell.

Markus Kneer

Success and Knowledge in Action: Saving Anscombe's Account of Intentionality

Abstract: According to Anscombe, acting intentionally entails knowledge in action. This thesis has been near-universally rejected due to a well-known counterexample by Davidson: a man intending to make ten legible carbon copies might not believe with confidence, and hence not know, that he will succeed. If he does, however, his action surely counts as intentional. Damaging as it seems, an even more powerful objection can be levelled against Anscombe: while acting, there is as yet no fact of the matter as to whether the agent will succeed. Since his belief that he will is not yet true while his action is in progress, he cannot possibly know that he is indeed bringing about the intended goal. Knowledge in action is not only unnecessary for intentional action, it seems, but–at least as regards success-bound types of action–impossible to attain in the first place.

In this paper I argue that traditional strategies to counter these objections are unsatisfactory and propose a new account of knowledge in action which has two core features: (i) It invokes an externalist conception of justification which not only meets Davidson's challenge, but also casts doubts on the tacit internalist premise on which his example relies. (ii) Drawing on recent work about future contingents by John MacFarlane, the proposed account conceives of claims to knowledge in action as assessment-sensitive so as to overcome the factivity objection. From a retrospective point of evaluation, previous claims about future events and actions can not only be deemed as having been true, but also as having been known.

1 Knowledge in Action

In order to count as acting intentionally, Anscombe (1963) claims, an agent must know what he is doing. More precisely, my φ-ing intentionally entails knowing

Note: This research was supported by the SNSF Ambizione grant PZ00P1_179912. Thanks to Barry Maguire and the participants of Prof. Peter Schaber's research colloquium at the University of Zurich for very helpful comments.

Markus Kneer, University of Zurich (markus.kneer@uzh.ch)

https://doi.org/10.1515/9783110702286-009

that I'm φ-ing; it is a necessary (though not sufficient) condition for intentional action. If, for instance, I am asked why I'm tapping my foot and respond that I wasn't aware of it, I do not count as having done so intentionally.

Despite its intuitive plausibility, the entailment thesis is near-universally rejected—primarily due to Davidson's well-known carbon copier counterexample. There is, however, an even more damaging objection that can be levelled against Anscombe, an objection which does not only question whether knowledge is a necessary condition for intentional action, but whether it is so much as possible. In the following, I will briefly outline Anscombe's theory of action in very broad strokes, introduce the objections and assess traditional strategies of defending the entailment thesis. Since the prospects of the latter are dim, I will sketch a new account of knowledge in action, which, I hope, can put these worries to rest. The resulting epistemology of action is less the fruit of Anscombe exegesis than an independent attempt to make sense of the entailment thesis. It is consistent with many of the rather idiosyncratic features of Anscombe's philosophy of action, yet relies on them as little as possible. As such, it is intended to appeal to both Anscombians and her critics alike.

1.1 Textbook Anscombe

The 'knowledge a man has when acting intentionally', Anscombe argues, is special in various respects. It is 'knowledge without observation', i.e. not based on perception or inference, and as such groundless.[1] "In opening the window," Anscombe writes, "I do not pause and think to myself: 'Let me see, what are my movements bringing about? The opening of a window.'" (1963: 51). To help intuition along, knowledge without observation is frequently compared to bodily awareness. Just as I know the position of my limbs without having to look, I do not have to observe my bodily actions in order to come to know what it is that I'm doing. The similarities between the two types of knowledge are limited though. As regards action, observational evidence *cannot* reveal what I am doing, since a particular episode of bodily behavior could count as any number of actions. When I'm tapping my foot, what is given through observation from a third-person perspective cannot settle whether I am communicating in codes with the tenant below, following the rhythm of the music or am engaged in

1 For recent discussion of Anscombe's theory of action, see Moran (2004) and Setiya (2007). For knowledge without observation in particular, see Pickard (2004) as well as the survey of the literature by Schwenkler (2012).

yet another action (cf. Anscombe 1963: 11). In order to be φ-ing intentionally, I have to conceive of my behavior under a particular description, that is, as φ-ing.[2] Importantly, what is known without observation is not merely what I am taking myself to do, or what I am trying to do, but—as Anscombe insists—*what happens*, namely, the event I am bringing about in the external world.

The epistemic attitude entailed by intentional action is thus distinctive in three respects: (i) *its content*, as the agent has to know what she is doing, and what is happening, under a particular description definitive of the action; (ii) its *source*, as the knowledge one has in acting intentionally is not derived from observation and finally (iii) *the character of the epistemic attitude*, which is rather demanding. It does not suffice to believe, or to believe justifiedly, that I am engaged in a particular action. According to Anscombe, nothing short of knowledge in the full-blown sense will do.[3]

Features (i) and (ii)—content and source—have given rise to considerable controversy. Regarding the third feature—the knowledge criterion—general opinion is more uniform: Although it is commonly accepted that intentional action is accompanied by some sort of psychological attitude, the claim that the latter amounts to an agent's *knowledge*, i.e. (at the very least) a true, justified, belief of what she is doing, has triggered widespread criticism. Such an account faces two principal problems. One concerns *true* belief, the other concerns *belief as such* (or, as I will argue, *justified* belief).

1.2 The Factivity Problem

Let's begin with truth. In φ-ing, I have no guarantee whatsoever that I will reach my goal. When attempting to swim across the Channel, for instance, I cannot be sure of success. I might well believe that I am swimming across the Channel, or I might know that I am attempting to swim across the Channel. But since there is, so far, no fact of the matter about whether or not I will succeed, the belief that I am *indeed* swimming across the Channel—that I am 'doing what happens'—cannot yet be true. Hence, there is no possible way for such a belief to amount to knowledge. In stubbornly *calling* this attitude knowledge, non-observational or

2 "[T]o say that a man knows he is doing X is to give a description of what he is doing under which he knows it" (1963: 12).

3 As is well known, there's a fourth feature: the 'practical', rather than 'speculative' nature of knowledge in action, in virtue of which what is known in action is 'the cause of what it understands'. I will turn to this peculiar characteristic as soon as it becomes relevant.

not, we would have to 'jump to conclusions' (Paul 2009) or take an epistemic 'leap of faith' (Langton 2004).[4]

In the literature, the fundamental nature of the factivity problem is not sufficiently appreciated and frequently glossed over.[5] For instance, those arguing against non-observational knowledge frequently end up with a 'two-factor approach' (a 'mad account' according to Anscombe): the agent knows what he intends non-observationally, yet his knowledge that he is in fact bringing about the intended event is based on perception.[6] Since there is not yet a fact as to whether the agent will indeed succeed in bringing about his goal while the action is still in progress, however, perception is an inadequate source of knowledge to solve the problem. The core problem concerns the *absence* of an object of knowledge rather than the appropriate *mode* to apprehend it.[7]

1.3 The Doxastic Problem

The second problem is that φ-ing intentionally might not even entail *believing* that one is φ-ing, as Davidson's famous carbon copier example (1978: 91–92) demonstrates. A man is attempting to make ten legible copies by pressing his pen hard on a stack of blank sheets interspersed with carbon paper. All the while, he is deeply sceptical of his success. As such, he does not believe that

---

4 Langton's primary target is not Anscombe's account, but Velleman's, which is even more demanding epistemically. Whereas Anscombe makes knowledge in action a necessary requirement for intentional action, Velleman (1989) identifies intention as such with a particular type of knowledge (wishful, self-fulfilling true belief).

5 A notable exception is Grice, who gives an early statement of the problem. His focus lies with future-directed intentions ('I intend to A'), but translates to as yet uncompleted actions ('I will indeed succeed in doing A'):

> A man who expresses an intention to do A (who says 'I intend to do A') is involved in a factual commitment; he is logically committed to subscribing, with this or that degree of firmness, to a factual statement to the effect that he will do A. [. . .] The standard source of entitlement to make such a factual statement is not available for this case, since the ordinary concept of intention is such that if one intends to do A, one is logically debarred from relying on evidence that one will in fact do A. No alternative source, however, of a different, non-evidential kind, for the entitlement to say 'I shall in fact do A' seems to be forthcoming

(1972: 8).

6 Cf. Adams & Mele (1989) and Falvey (2000) for discussion.

7 This is not to say that the epistemic reach of non-observational knowledge should not be scrutinized. It might prove inadequate for reasons independent of factivity. We should be careful, however, not to level the factivity problem against non-observational knowledge *qua* non-observational knowledge, as is not uncommon.

he will indeed produce all the ten copies he needs. If he succeeds, however, it would be astonishing to deem his action non-intentional. But if φ-ing intentionally does not entail that the agent believes that he is φ-ing, it will certainly not entail that he knows he is φ-ing. Let's call this the 'doxastic problem' for knowledge in action. Given that knowledge in action seems neither necessary (the doxastic problem) nor even so much as possible (the factivity problem), Anscombe's account might sound completely implausible.

In the next section, I will briefly survey three attempts to defend Anscombe's proposal, all deemed insufficient for one reason or another. Thereafter, I'll argue against the two objections just raised. The goal is not to defend Anscombe's comprehensive action theory in word and letter, but rather to show that the plausibility of the entailment thesis is not as easily dispelled as is frequently assumed. I will thus largely refrain from a cumbersome exegesis of Anscombe, yet occasionally point out how the suggested account squares with the more general picture proposed in *Intention* (1963).

2 Defending the Knowledge Criterion

2.1 Reduction in Scope

There are different types of strategies to defend some variation of the knowledge criterion. One might, for instance, attempt to reduce its scope by imposing restrictions on what it is that needs to be known by the agent. Even in the carbon copier example, there is *something* the agent knows about his actions and, it might be argued, it is in virtue of that knowledge that the action counts as intentional. As Davidson himself acknowledges, when acting intentionally, what the agent does is "known to him under some description" (1971: 50). The carbon copier, we might hold, certainly knows that he intends to make ten copies (Donnellan 1963), or that he is trying to make ten copies (O'Shaughnessy 2003).

Some authors deem such a change in the object of knowledge inappropriate. They propose to stick to the object of knowledge envisioned by Anscombe, which does not consist in the agent's intending or his trying to do something, but his very doing and that which happens. The complications are to be accommodated by opting for a less demanding epistemic attitude, that is to say, belief rather than knowledge. Yet others hold that the only way to defend some form of the knowledge criterion is to opt both for a different epistemic attitude and a more modest object of knowledge. Setiya (2007), for instance, does precisely that. He proposes that the agent must act in the belief of doing particular things such as, in Davidson's example, pressing hard on the paper, with the end of mak-

ing ten copies. On this account, it suffices that the agent believes, rather than knows, that he is engaging in some of the actions constitutive of bringing about his more general goal. Such strategies are, as I attempt to show below, unsatisfactory because they concede too much.

2.2 Action Descriptions in Progressive Aspect

A more promising path to pursue is the appeal to the 'broadness' of the progressive of action verbs.[8] Action descriptions involving atelic verbs in progressive form ('is swimming', 'is walking' etc.) do not have a success condition built in and thus impose limited epistemic demands. While swimming, I can know at any point in the course of doing so that what I am taking myself to do is—hard-core scepticism aside—in fact happening. No epistemic leaps of faith required. In the case of atelic action verbs, the truth of a description in the progressive tense licenses the truth of its description in the past tense: if I am swimming, it will be the case that I swam or that I have been swimming. For telic action verbs—verbs with a success condition built into their semantics—that is not necessarily the case: if I am trying to swim across the Channel, and turn around after a few strokes, it will not be true that I have crossed the Channel or that I was crossing the Channel. While trying to cross the Channel, it is simply impossible for me to know that I am indeed getting to the other side since any such fact has not yet been established.

The flexibility introduced by the progressive tense, however, is not limited to atelic action descriptions. It extends to a considerable class of telic actions—we might call them 'weakly telic'—even if the stipulated goal rests unfulfilled. For instance, one might count as crossing the street though one never makes it to the other side.[9] Actual completion is not essential—nothing stands in the way of granting the agent knowledge of her actions, just as in the case of atelic action descriptions. However, in many cases (call them 'strongly telic') the completion of the goal stipulated by the action verb is essential for the action to be considered as taking place or for the event to happen. I don't count as crossing the

8 Cf. Anscombe (1963: 39), developed in Falvey (2000). Further discussion in Thompson (2008, Ch. 8), Paul (2009), Haddock (2011) and Schwenkler (2012).

9 Hence Anscombe's insistence that "a man can be doing something which he nevertheless does not do" (1963: 39). The interesting though somewhat murky distinction between weakly and strongly telic actions extends to the past progressive tense. Whereas it sounds infelicitous to say 'Mary was crossing the Channel' if she returned ashore after a few strokes, there's nothing wrong with saying that Frank 'was crossing the street' if he turns back or gets run over.

Channel if I give up after a few strokes, as finishing my soup or killing a fly if I don't, or as walking towards the library if the distance between it and me is not decreasing.[10] But if strongly telic actions are success-bound, how can I know that I am indeed doing what I intend to do? The appeal to the broadness of the progressive tense is not effective here to overcome the factivity problem. As such, it can only serve as a partial defence of the entailment thesis.

2.3 Practical Knowledge

The third strategy to defend the knowledge criterion takes its cue from yet another peculiarity Anscombe has in stock when it comes to knowledge in action. Such knowledge, she proposes, is not like ordinary, "speculative" knowledge, i.e. passive or "receptive," in so far as it aims to fit the facts or is "derived from the objects known." Rather, it is "practical" in nature and, as such, "the cause of what it understands" (Anscombe 1963: 87). If I happen to be mistaken as to what I'm doing, "the mistake here is one of performance, not of judgement" (5, 56, 57, 87–89). The precise nature of practical knowledge remains one of the more elusive chapters of Anscombe's theory of action. Strategically, this way to save the entailment thesis should nonetheless already be obvious: it consists in cashing in maximally on the resources provided by practical knowledge so as to help overcome the factivity problem. However, the more such practical knowledge is moulded into a type of epistemic state that does not constitutively aim at representing the facts, the less it deserves its name and the more bewildering the resulting picture of intentional action.[11]

This point can be made in somewhat more detail. Let's have a look at Anscombe's discussion of the difference between speculative (or 'contemplative') and practical knowledge:

> Can it be that there is something that modern philosophy has blankly misunderstood: namely what ancient and medieval philosophers meant by practical knowledge? Certainly in modern philosophy we have an incorrigibly contemplative conception of knowledge.

10 Last example by Haddock (2011). Here the strongly telic aspect seems to be imposed not by the verb but by the preposition. Interestingly, there is not only a general success condition at work in this type of action description, but a stipulated manner of how the result is to be brought about. Though rather different from standard strongly telic actions, they pose a similar problem for the entailment thesis.

11 For a discussion of Anscombe on practical knowledge, cf. Velleman (1989), Moran (2004), Setiya (2007, 2008, 2009), McDowell (2011), Haddock (2011), Thompson (2011) and Schwenkler (2011).

> Knowledge must be something that is judged as such by being in accordance with the facts. The facts, reality, are prior, and dictate what is to be said, if it is knowledge. And this is the explanation of the utter darkness in which we found ourselves. For if there are two knowledges—one by observation, the other in intention—then it looks as if there must be two objects of knowledge; but if one says the objects are the same, one looks hopelessly for the different mode of contemplative knowledge in acting, as if there were a very queer and special sort of seeing eye in the middle of the acting (1963, 57).[12]

Here's one way to interpret this passage: the contemplative conception of knowledge seems unsuited for knowing what one is doing, since "facts, reality, are prior, and dictate what is to be said"—and presumably also what can be known. The factivity of contemplative knowledge would require "a very queer and special sort of seeing eye in the middle of the acting"—an eye that anticipates what will come to pass, an eye that jumps to conclusions. Introducing a second type of knowledge of a practical nature seems to bring an extra complication into the picture: if there are two types of knowledge, should there not be two different objects of knowledge as well? Practical knowledge, we might think, captures what I am (or take myself to be) doing, contemplative knowledge what happens. But multiplying the objects of knowledge would be a mistake, as Anscombe vigorously argues in various places. If things work out, the action and the event are one and the same thing—'I do what happens'. In virtue of being defined by the same description, they must constitute a single object of knowledge.[13] Sticking with a single object of knowledge, however, does not prove the futility of a second kind of knowledge: a single object of knowledge might be approached via different modes of knowledge, a ball can be known to be spherical through touch or vision. Hence the question whether 'modern philosophy', with

12 As so often happens in *Intention*, various strands of the discussion are run together. From this passage it seems that the dichotomy between observational and non-observational knowledge maps onto the one between speculative and practical knowledge. This is of course not the case. A priori knowledge, such as mathematical knowledge, is clearly responsive to facts in ordinary ways without drawing on observational input.

13 Here's a characteristic passage: "The difficulty however is this: What can opening the window be except making such-and-such movements with such-and-such a result? And in that case what can *knowing* one is opening the window be except knowing that that is taking place? Now if there are two *ways* of knowing here, one of which I call knowledge of one's intentional action and the other of which I call knowledge by observation of what takes place, then must there not be two *objects* of knowledge? How can one speak of two different knowledges of *exactly* the same thing? It is not that there are two descriptions of the same thing, both of which are known, as when one knows that something is red and that it is coloured; no, here the description, opening the window, is identical, whether it is known by observation or by its being one's intentional action" (1963: 51).

its preference for a contemplative conception of knowledge is not up a blind alley and should rather approach the same object of knowledge via a different epistemic mode. What epistemic mode? Practical knowledge. And practical knowledge, to repeat, is not "derived from the objects known", but the "cause of what it understands".

One might reject Anscombe's account of practical knowledge outright as "causally perverse and epistemically mysterious" (Velleman 1989: 103). As regards the alleged causality, however, we could follow Moran's suggestion and treat practical knowledge as the formal rather than the efficient cause of what it understands. In virtue of the operative description which defines my doings as a particular action, practical knowledge specifies intensionally, rather than causes extensionally what it understands. Anscombe's slogan "concerns the formal or constitutive role of the description embedded in one's practical knowledge making it the case that this description counts as a description of the person's intentional action. If the agent didn't *know* this happening under *this* description, then as so specified, it *would not be* 'what he is intentionally doing'" (Moran 2004: 54, italics in the original).

Though the charge of causal perversity might thus be averted, its potentially non-factive nature remains a mystery—a mystery that is captured well, yet not resolved, by Anscombe herself:

> If then my knowledge is independent of what actually happens, how can it be knowledge of what does happen? Someone might say that it was a funny sort of knowledge that was still knowledge even though what it was knowledge of was not the case! On the other hand Theophrastus' remark holds good: 'the mistake is in the performance, not in the judgment' (1963: 82).

Since a non-factive conception of practical knowledge would be a "funny sort of knowledge" indeed, this strategy to save the entailment thesis might be deemed heroic, but remains implausible from the outset. Theophrastus' dictum does little to dispel the worries.

Let's briefly take stock. On the one hand, not knowing I was φ-ing is an excellent explanation of my not φ-ing intentionally (subconscious conundrum aside). On the other hand, it is difficult to explain how it is even so much as possible to know that I am actually φ-ing while φ-ing, at least as long as we refuse to make do with atelic actions only, water down knowledge to knowledge of trying or mere belief and shy away from an utterly implausible, non-factive conception of 'practical' knowledge. But the knowledge criterion can be defended, I think, without making any concessions. In the following, I'll attempt to show how.

3 Knowledge Proper and the Factivity Problem

Intentional action, conceived as entailing knowledge of what one is doing, we have said, gives rise to two fundamental problems. Firstly, as Davidson has argued, in acting intentionally, one might not need to hold a corresponding belief, let alone know what one is doing. Secondly, it would be reassuring if we could leave epistemology proper in place, i.e. agree that "[k]nowledge must be something that is judged as such by being in accordance with the facts"—but it is not obvious how this could be so much as possible. Both problems are strongly related to the success of one's action: in the former case, what undermines belief is a low perceived probability as to one's abilities to succeed. In the latter case, prior to having successfully completed one's action, there simply is no fact that can be known in acting intentionally.

I will argue that neither problem is specific to action theory. The very possibility of knowledge regarding one's action is a variation of the semantic puzzle of future contingents, a puzzle that can be solved. Davidson's counterargument, I will suggest thereafter, is less a matter of belief than of justification. As such it raises questions regarding internalist and externalist accounts of epistemic justification, and thus concerns a complication of epistemology proper rather than one of action theory narrowly conceived. Whereas the general debate between internalists and externalists has not given rise to a consensus after decades of argument, externalism, I'll suggest, could well carry the day as regards knowledge in action.

3.1 Assessment-sensitivity of Future Contingents

Let's begin with the factivity problem. On an indeterminist view, the future course of affairs is a contingent, rather than a settled, or necessary, matter. At every moment in time, there is a variety of genuine possibilities as to what world will be actual in the future. Along which branch the world will develop as the future unfolds is more than just epistemically indeterminate, it is metaphysically indeterminate as there is as yet no fact of the matter.

Indeterminism gives rise to a semantic paradox regarding utterances of future contingents. Take the following example:

(1) Frank: "It will be sunny tomorrow."

At Frank's context of utterance, whether it will be sunny the next day is still indeterminate. Accordingly, the proposition[14] expressed by (1) must be judged as neither true nor false. Call this, following MacFarlane, the *indeterminacy intuition*. The next day, let's assume, the sun is shining. In retrospect, we are inclined to hold that what Frank said was true. That is, we deem the proposition expressed by Frank's utterance as true at his context of utterance. Call this the *determinacy intuition*. Problematically, orthodox semantics cannot accommodate these conflicting intuitions.

In Kaplanian semantics[15] a *context* is a potential occasion of utterance for a sentence S which plays a twofold role. On the one hand, it determines the *semantic values* for the indexical expressions of the sentence. In Kaplan's terms, the context and the character of the sentence (or sentence type) jointly deliver the content expressed. Though your utterance 'I am cold' and mine have the same character, they express different contents because the context provides different individuals as the semantic value of the indexical 'I'. On the other hand, the context determines the *circumstances of evaluation*. That is, it specifies under which set of parameters (a world, a time, potentially a location, a standard of precision etc.) the truth or falsity of the content should be evaluated. If on Monday, drenched to the bones, I utter 'It's raining' and claim the same the next day in bright sunshine, the difference in truth value is accounted for by the fact that my different contexts of utterance specify different time parameters and hence different circumstances of evaluation. In the first case, the circumstance of evaluation is the actual world on Monday, in the second it's the actual world on Tuesday. Though identical in content, my utterance on Monday is true, but my utterance on Tuesday is false. Importantly, whether or not my utterance was true yesterday depends on the context of utterance, not the context of assessment—there is no such thing on this view. On Tuesday, yesterday's claim that it was raining *remains* true, because the time parameter in the circumstances of evaluation provided by the circumstance of utterance is Monday. The only context that affects the circumstances of evaluation is the context of utterance, so there cannot be any truth value switching which might explain our contradictory intuitions regarding future contingents.

14 This section draws heavily on the work of John MacFarlane. His original statement of the paradox was phrased in terms of *utterances* (MacFarlane 2003), but neither the puzzle nor its solution resist formulation in terms of uttered propositions (MacFarlane 2008, 2014). For a critical discussion cf., for instance, Cappelen & Hawthorne (2009) and the contributions to *Around the Tree* (Correia & Iacona 2012).

15 Cf. Kaplan (1979, 1989).

A standard way to make sense of the indeterminacy intuition is to seek recourse to supervaluationism, a semantics that besides 'true' and 'false' introduces a third value, 'indeterminate' (or 'neither true nor false'). According to Thomason (1970), for instance, an utterance is true or false *simpliciter* if true or false in all possible worlds and otherwise indeterminate. Frank's claim 'It will be sunny tomorrow' is evaluated as neither true nor false, and even though the next day the sun happens to be shining, it is still not true in all possible worlds, therefore this evaluation won't change. Though indeterminacy is accounted for, our retrospective determinacy intuition is not.

MacFarlane's solution[16] to the puzzle is to expand the roles contexts can play: besides the context of utterance, we also have to take the context of assessment into account. Presume that on Monday (t_1) there is a genuine possibility for it to rain the next day (t_2). This means that the world actual at c_1 branches into sunny worlds (w_1) and rainy worlds (w_2, w_3), as illustrated in Figure 1. Which of these worlds will turn out actual is, on Monday, still not settled. The context of Frank's utterance, c_1, specifies a circumstance of evaluation comprising of the world of utterance (up to then $w_@=w_1=w_2=w_3$) and the current time t_1. At that circumstance, Frank's claim 'It will be sunny tomorrow' is indeterminate. Once the future has unfolded and w_1, a sunny world, has turned out actual, (1)—as uttered at c_1 ($w_1=w_2=w_3$, t_1) and *assessed* from c_2 (w_1, t_2)— is true.

According to the supervaluationist picture, retrospective evaluation of (1) takes into account all worlds overlapping at the context of utterance c_1, that is, w_1, w_2 and w_3. In virtue of not being true at all these worlds, (1) is, and remains, indeterminate. On MacFarlane's view, the retrospective evaluation of (1) focuses on only *some* of the worlds overlapping at the context of utterance, namely those which *also* branch through the context of assessment. If it turns out that the latter is c_2, (1) was true as uttered at c_1 and assessed from c_2; if it turns out that the context of assessment is c_3 (w_3, t_2), (1) was false as uttered at c_1 and assessed from c_3. Both the intuition of prospective indeterminacy and retrospective determinacy are borne out.

16 I am simplifying considerably. What matter for our purposes is the basic idea. For details regarding the semantic framework, see MacFarlane (2014).

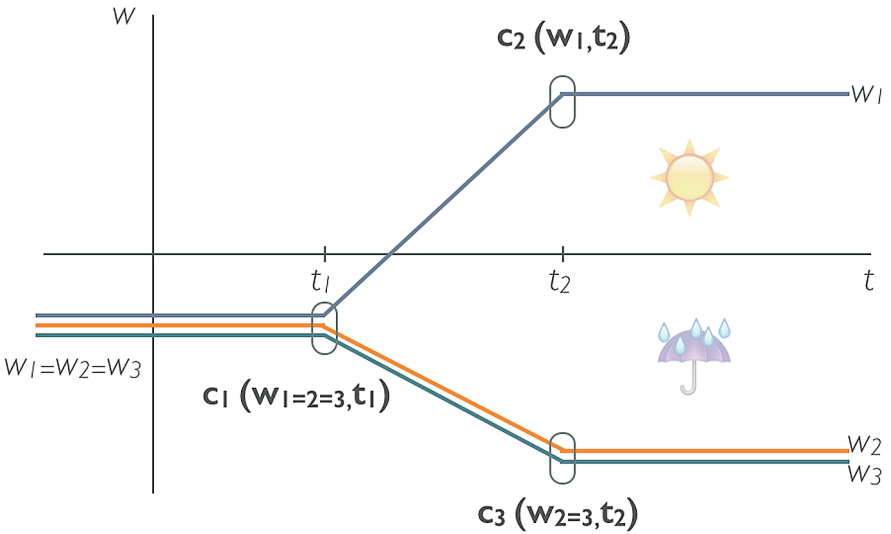

Figure 1: Context-sensitivity in an indeterministic universe.

3.2 Assessment-sensitivity of Knowledge Attributions

The curiosity regarding the *truth* of a proposition uttered or entertained extends, I think, to the attribution of *knowledge* of propositions entertained. Conceive of knowledge as a justified, true belief, and assume that Frank's belief that it will be sunny tomorrow is well justified. In an indeterministic universe, Frank cannot know today that it will be sunny tomorrow, because his belief is not true yet—rain tomorrow is a genuine metaphysical possibility. In retrospect, however, if Frank was right and had good reasons for his belief, it is perfectly felicitous to say that Frank *knew* it would be sunny today. With hindsight, not only the alethic assessment becomes more determinate (a truth-indeterminate proposition gets a determinate truth value) but also the assessment of the subject's epistemic state (a truth-indeterminate justified belief becomes a true justified belief, i.e. knowledge, or a justified false belief). At a context of assessment which equals the context at which Frank entertains his belief, his epistemic state is one of justified belief. Once the future has unfolded and w_1, a sunny world, has turned out actual, claim (1), as entertained at c_1 (w_1=w_2=w_3, t_1) and *assessed* from c_2 (w_1, t_2), will

be true and so Frank will count has having held a true, justified belief—that is, he will count as *having known* that it would be sunny the next day.[17]

The fact that knowledge attributions are context-sensitive is not news.[18] For instance, in standard scenarios, John might count as knowing that his car is in his driveway, if that's where he left it. If someone voices the possibility that it might have been stolen, however, John will likely concede that he doesn't know. The revision is less a matter of correcting a *mistaken* claim to knowledge, than one of adapting to a different, more demanding standard of knowledge. John will not only acknowledge that he now *doesn't* know, but also that he *didn't* previously know the whereabouts of his car before the standards were raised. To make sense of this retraction, we have to evaluate John's claim not according to the (low) epistemic standards of the context of utterance but according to the (higher) standards of the context of assessment. The context-sensitivity at play in such cases, however, is importantly different from the one developed above. What explains the variation in knowledge attributions when it comes to different epistemic standards are variations in *justification*. According to ordinary standards, John's belief that his car is in the driveway is justified. Judged by a different—higher—set of standards, it might not be, and the belief will thus not count as knowledge. What explains the variation in knowledge attribution in the previous paragraph, by contrast, regards not justification but the *truth* of the subject's belief. Though the context-sensitivity thus affects different constituents of knowledge (truth v. justification), they manifest the same basic relativist logic: knowledge ascriptions, it seems, are sensitive not only to contexts of entertainment/utterance but to the context of assessment in several ways.[19]

17 Attacking the premise of an open future does not weaken the argument, but makes it, if anything, stronger. In that case knowledge ascriptions at the context of utterance are no longer problematic due to metaphysical and epistemic indeterminism, but only due to the latter: In a deterministic universe, the future course of events is settled and beliefs about it are either true or false. Naturally, on certain accounts of justification we might still be loath to call such beliefs knowledge before the future has materialized. From a suitable context of assessment, however, future contingents can not only be understood as having been true (obviously so, due to determinism), but—if well justified—as having been known.

18 See, for instance, MacFarlane (2005), who covers context-sensitivity regarding standards of precision. The example discussed in the main text is his.

19 The application of assessment-sensitivity to epistemic expressions has become a fertile field of research recently. Besides the mentioned standards of precision, epistemic modals can also be seen as sensitive to contexts of assessment (cf. Egan, Hawthorne & Weatherson 2005; Egan 2007; Stephenson 2007, MacFarlane 2011). The linguistic intuitions inspiring such accounts, however, are not uncontroversial (cf. Hawthorne 2007, Von Fintel & Gillies 2007, 2008; Yalcin 2011; Braun 2012; Kneer 2015, 2020). Still, one can safely accept the assessment-sensitivity of future contingents—where intuitions are considerably more uniform—without buying into a rampant relativ-

3.3 Assessment-sensitivity of Knowledge in Action

Utterances and beliefs regarding actions underway are utterances and beliefs regarding events in progress. If the latter can be true, and known, at a context of being in progress when assessed from a context of completion, so can the former. There is thus nothing mysterious about the "knowledge a man has when acting intentionally". From a retrospective context of successful completion, the agent, if holding a suitably justified belief, can be understood as *having known* that she was acting as intended at the context of action in progress. There is no reason to presume that she is making an epistemic leap of faith, or jumping to conclusions. This is of course perfectly in line with our ordinary ways of speaking and acting. Having completed an action successfully while having had good justification that I would, it's as natural for me to say that I knew I was baking a cake or that I knew I was making the boss uncomfortable as it is to say that I knew there'd be a department meeting today. The opposite would be deeply counterintuitive: if I can have knowledge regarding future events, it would be astonishing if I could not have knowledge regarding future events which I can directly influence and whose progress I can monitor.

Just as the attributions of intentionality and knowledge in action go hand in hand on this account if the action is successful, the absence of one feature will coincide with the absence of the other.[20] When my action is unsuccessful, my claim to practical knowledge fails. My belief about my doings, despite being maybe well justified, turns out false. Since knowledge is factive, I simply cannot have known. Relatedly, if I unintentionally bring about B, while trying to do A, I also lack knowledge in action. I thought of myself as bringing about A, not B, and as long as B is not constitutive of A, I will cite precisely this fact—that I didn't know I was doing B—as evidence in favor of the assessment that I did not do B intentionally.

The account of knowledge in action developed here leaves the fundamental traits of Anscombe's picture in place. The relevant type of knowledge is still practical in Moran's qualified sense and differs from purely 'speculative' knowledge. My conception or 'description' of what I am doing is an essential constituent of the knowledge I have when acting intentionally. It defines my behavior as a particular action, and it is in this sense that practical knowledge should be con-

ism regarding epistemic notions in general. Besides, while the solution sketched here is developed in terms of MacFarlane's relativist semantics, a similarly strong case could be made by employing more conventional frameworks such as, for instance, the one developed in Brogaard (2008).

20 I am ignoring counterexamples *à la* Davidson for the moment, but will turn to them shortly.

ceived as "the [non-efficient] cause of what it understands". Knowledge in action so understood is also in some fundamental way non-observational: I do not pause and look to see what I'm bringing about. Given the underdetermination of observational evidence as regards the definition of an action, it's simply impossible to find out. It is in virtue of the very description I have of my doings that they count as the action in progress, and that description I surely know without observation.

The question whether I can know what I am doing, and what happens, *exclusively* in a non-observational fashion remains, of course, highly controversial. Moran, for instance, argues that practical knowledge in action, despite being in an important sense non-observational, is nonetheless aided by observation. I do not want to get entangled in this debate. One brief point, however: the question of non-observational purity can be rephrased as the question whether we can only attribute knowledge in action *post-hoc*, if the agent has made sure by perceptual means that he indeed fulfilled his aim. *Prima facie*, such a condition does not seem necessary, in which case the chances for exclusively non-observational practical knowledge might stand better than frequently assumed.

4 The Doxastic Problem

4.1 Belief and Justification

Davidson's famous counterexample is frequently reported thus: the carbon copier, skeptical as to his success of producing ten legible copies, cannot be said to know that he will make ten copies, because he does in fact not even believe that he will. This argument, according to which knowledge is *a fortiori* out of the question, is rather unconvincing and it is not what Davidson had in mind.[21] In clear-cut cases in which an agent lacks the belief that he is φ-ing, his φ-ing will not be deemed intentional. For instance, if a man scribbles absentmindedly on a stack of carbon papers (hence lacking the belief that he is) and produces ten legible copies, there is little doubt that he didn't make them intentionally. Or, to take another type of case, presume someone attributes zero probability to his succeeding in φ-ing, in which case he cannot be said to believe that he is φ-ing. Pursuing his action nonetheless is much rather a manifestation of *irrational*

21 Things are of course different if we follow Williamson's (2000) suggestion according to which the verb 'to know' picks out a *sui generis* mental state that cannot be factored out into more basic constituents.

behavior than of intentional action, quite independently of whether the goal is attained. The entailment thesis is not under pressure from such cases.

Now, beliefs do not have to be held with certainty. If an agent attributes a low probability to the truth of his belief, it is no less a belief—even though, on some accounts, it might be deemed unjustified. I may, for instance, cling to the belief that it will be sunny in London all summer despite being aware that the probability of that is extremely low. Of course, if it turns out to be sunny in London all summer, my belief, in virtue of being poorly justified, will not amount to knowledge. The carbon copier—call him Donald—who considers his chances of success relatively low, but not inexistent, can thus be conceived as believing to some degree, but arguably not as knowing, that he is making ten copies. The doxastic problem, as we called it, does not derive from an *absence* of belief, but rather from a lack of epistemic justification for the belief regarding one's action.

4.2 Internalist and Externalist Justifications

Let's take a look at how Davidson phrases his example:

> [I]n writing heavily on this page I may be intending to produce ten legible carbon copies. I do not know, or believe with any confidence, that I am succeeding. But if I am producing ten legible carbon copies, I am certainly doing it intentionally (Davidson 1978: 2001: 80).[22]

The argument thus comes to this:

1. I can φ intentionally without believing with confidence that I'm indeed φ-ing.
2. Knowing that I'm φ-ing entails believing with confidence that I'm φ-ing.
3. Therefore, I can φ intentionally without knowing that I'm φ-ing.

The crucial premise is of course the tacit one, i.e. (2). What is the philosophical motivation for such an assumption? The only explanation that comes to mind is that Davidson implicitly pledges allegiance to (a strong form of) epistemic internalism. Internalism is the view that one must be aware of, or be able to become aware of, the grounds that justify one's belief that p. Differently put, justifiers must be 'internal' to and directly accessible from the subject's cognitive perspective. On another account, the subject has to fulfill certain epistemic duties in order to count as holding a justified belief. Attributing a low subjective probabil-

22 For a slightly different formulation, cf. also Davidson (1971, in 2001: 50).

ity to the belief that one will actually φ can be seen as a *defeater* to the internalist belief justification. Lacking justification, the belief is no longer a candidate for knowledge. The man who is skeptical about whether he will succeed in making ten legible copies does not hold the belief that he will indeed succeed with justification and can thus not be considered knowing that he will.

The alternative to epistemic internalism is externalism.[23] Weak externalism denies that justification should exclusively be conceived as an internal matter, a strong form denies that justification is ever internalist. According to this view, the justifiers of a belief need not be accessible to the believer, as they can be external to her cognitive perspective. What matters instead is (according to one approach) that the subject's belief *reliably tracks the truth*, whether or not she thinks it does. An unconfident examinee reliably giving the right answer to a certain question can be seen as knowing, despite the fact that she is unsure and has no inkling of how she came about the information.

Adapted to knowledge in action, we might similarly consider self-confidence or subjective probability less important than objective probability, that is, the agent's reliable disposition to successfully carry out the intended action. In this case, the justification of my belief that I'm φ-ing will not be defeated by my lack of self-confidence; what justifies my belief can be external to (and hence unimpinged by) my cognitive perspective. According to such a view, there is no *prima facie* problem in attributing knowledge to Davidson's carbon copier. The man's reliable disposition to make ten copies (whether or not he has confidence in himself) is sufficient to justify his belief and thus to warrant the ascription of knowledge. If, on the other hand, the man's objective probability of success is extremely low, his belief that he will make ten copies should not count as knowledge.[24]

4.3 Objective and subjective probability

Let's bring the proposed view into somewhat sharper focus by aid of an example. Sitting in front of a stack of 100 sheets interspersed with carbon papers, Mary is

23 Chisholm (1966), Bonjour (1985) and Lehrer (2000) are well-known advocates of internalism; Goldman (1967), Armstrong (1973), Dretske (1971, 1981) and Nozick (1981) are externalists.

24 The variations of internalist and externalist approaches to justification are plentiful and the line between the two types of account can be drawn in very different ways (cf. Kim 1993). The internalist/externalist dichotomy, developed here in terms of subjective and objective probability, is orthogonal to some of the possible distinctions, yet fits the general thrust of several accounts naturally enough.

trying to make exactly 57 copies—not one more, and not one less. Like Donald, she deems the probability of success low. If she, to her surprise, succeeds, we would call this a lucky accident rather than an intentional action. Superficially, the two scenarios are very similar: In both cases the agents attribute a low probability of truth to the belief that they are indeed doing what they intend to do. Due to the perceived marginal chances of success, the beliefs are poorly justified—neither agent has good grounds to believe they will fulfil their goal. Both beliefs turn out true. We cannot attribute knowledge to either of the agents, yet Donald's action counts as intentional whereas Mary's as a fluke. Why is that?

The difference between the two cases is, of course, one of objective probability of success.[25] In Mary's case, the probability is extremely low, in Donald's it is considerably higher. Whether or not an action counts as intentional, it thus seems, depends less on the subjectively perceived probability of success, and more on objective probability, i.e. on the actual difficulty of the task for the agent. This gives us a matrix of four basic cases (Table 1): Let's start with Donald and Mary, who are both skeptical about their success (bottom row of the table), but Donald's task is conceived as comparatively easy, whereas Mary's as hard. Intuitively, if successful, Mary's action counts as a lucky accident, whereas Donald's counts as intentional.[26] Presume that John also wants to make ten carbon copies and that Sally also wants to make 57 out of 100 possible ones, but in comparison to Donald and Mary they are both confident as to their success (upper row of the table), i.e. they attribute a high subjective probability to their belief's being true. Table 1 summarizes the different cases for the success condition:

Table 1: Dependence of intentionality and knowledge on objective and subjective probability.

	High objective probability	**Low objective probability**
High subjective probability	*John* Intentional Action Knowledge	*Sally* Lucky Accident ?
Low subjective probability	*Donald* Intentional Action ?	*Mary* Lucky Accident No Knowledge

25 More precisely, what matters is the estimated objective probability of success as perceived from a perspective external to the subject—the same perspective from which the intentionality of the action is assessed. It is not the (frequently unknown) actual objective probability of success that matters, because the latter is not generally what informs our ascriptions of intentionality or our assessment of an agent's epistemic situation.

26 In fact, empirical evidence suggests that intentionality ascriptions might depend only on objective probability and not on subjective probability. See, for instance, Mele & Cushman (2007).

Where subjective and objective probability coincide, the assessment is uncontroversial: John's action is intentional, his belief is both true and justified (no matter whether we're internalists or externalists) and hence an instance of knowledge. Mary, if successful, would be deemed lucky. Due to the low probability of success, of which she is well aware, she cannot justifiably claim that she knew she was indeed making 57 out of 100 copies. Such cases of homogenous probabilities are consistent with both Anscombe's and Davidson's views.

Things get complicated where the two types of probability diverge. To attack knowledge as a necessary condition for intentional action, Davidsonians argue that Donald doesn't have knowledge in action, whereas his action, if successful, is nonetheless judged intentional. Considering his chances of success low, Donald does not have reasons to take his belief— i.e. that he will indeed make ten copies—to be well justified. If said belief, unjustified from his perspective, turns out true, it would, according to this picture, come somewhat as a surprise to hear him insist that he knew all along that he was making ten copies.

From an external perspective, however, the belief held by Donald with a low degree of confidence is not unjustified: there is a high enough objective probability of succeeding in simple tasks such as this one. From such an external perspective, Sally's belief that she will make 57 out of 100 copies, despite being held with great confidence, on the other hand, does not seem well justified. Hence, on such an externalist account of justification, Donald can be said to know what he's doing, whereas Sally cannot. Differently put, according to such a view, Anscombe's entailment thesis faces no obstacle.[27]

---

27 The general recipe for construing counterexamples to the entailment thesis consists in maximizing objective probability by invoking an extremely simple everyday action, while minimizing subjective probability through imbuing the agent with legitimate doubts as to the success of his actions. Here's a well-known scenario, standardly attributed to Bratman:

> I'm recovering from paralysis of my right hand. I try to clench my fist, though am sceptical whether I'll succeed. If I succeed, I cannot be said to have known I would. However, I have surely clenched my fist intentionally.

The simplicity of the action guarantees that it is assessed as intentional. The legitimacy of my doubts prevent us from ascribing knowledge. The trick lies not only in stipulating a strong mismatch between the two types of probabilities, but in obfuscating the fact that their contrasting levels should affect each other and hence the attribution of knowledge and intentionality: if I do have legitimate doubts as to whether I can indeed clench my fist—something that generally takes months of practice after paralysis—then it is by no means obvious why my unlikely success should count as intentional. An objective probability of five percent does not make clenching my fist more intentional than hitting the bull's eye with bow and arrow, just because the former is standardly conceived as an easy task and the latter is not. If, on the other hand, I am indeed

Objective probability captures whether a certain action intended by an agent reliably comes to pass. As such, it is objective probability, not the agent's confidence, that determines whether we deem an action intentional or not. It is suggested that the same type of probability also determines whether we are willing to consider an agent's belief about his action in progress justified.[28] The strong relation between intentionality and knowledge in action postulated by Anscombe can thus at least partially be accounted for by the fact that both intentionality and belief justification derive from a common source—objective probability—and thus give rise to the same necessary condition. If I am able to φ reliably, and my φ-ing counts as intentional in virtue of that disposition, then what matters for justifying my belief that I'm φ-ing similarly seems to be whether I am able to φ reliably, not whether I so *consider* myself. Correspondingly, my inability to φ reliably not only undermines my claim that I was φ-ing intentionally, but also the justification of my belief that I did.

Davidson's insistence on confidence, on the other hand, is not easy to make sense of. What subjective probability tracks, if things go well, is precisely one's ability to bring about the intended action, that is, the objective likelihood of success. In the overwhelming majority of cases where the tracking works well, it is a very useful device. When it doesn't, it seems more appropriate to look to objective probability itself for justification rather than to its poorly calibrated proxy. Knowledge in action, misconceived in a way in which said proxy does play the central justificatory role, is—in line with what Davidson holds—most certainly not a necessary condition for intentional action.

In sum, Davidson's scenario, despite its clever trading on disjoint subjective and objective probabilities, does little to cast any doubt on Anscombe's epistemology of action. An externalist account of justification makes the entailment thesis not only possible, but is also considerably more plausible than an internalist approach to knowledge in action.

---

healed and have the usual (or similar) objective chances of clenching my fist, it's not at all clear why my pessimist attitude should be of any major epistemic consequence. The action is so simple that it is hard to believe how anyone with standard chances could *not know* that they're clenching their fist when they are, whether they are confident or not. Needless to say, if we buy into the logic of these examples, we'll have to deny all people with a naturally pessimist or skeptical disposition knowledge of much of what they are doing.

28 Preliminary experimental results confirm that 'folk' ascriptions of justification and knowledge do in fact correlate with objective probability and do not correlate with subjective probability (cf. Kneer, in prep.).

References

Adams, Frederick and Mele Alfred. 1989. “The role of intention in intentional action”. *Canadian Journal of Philosophy* 19(4): 511–531.

Anscombe, G. E. M. 1963. *Intention*. Oxford: Blackwell Press, 2nd edition edition.

Armstrong, D. M. 1973. *Belief, Truth and Knowledge*. Cambridge: Cambridge University Press.

BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge: Cambridge University Press.

Braun, David. 2012. “An Invariantist Theory of ’Might’ Might Be Right”. *Linguistics and philosophy* 35(6): 461–489.

Brogaard, Berit. 2008. Sea Battle Semantics. *The Philosophical Quarterly* 58(231): 326–335.

Cappelen, Herman and Hawthorne John. 2009. *Relativism and Monadic Truth*. Oxford: Oxford University Press.

Chisholm, Roderick M. 1966. *Theory of Knowledge*. Englewood Cliffs, (N.J): Prentice Hall.

Correia, Fabrice and Iacona Andrea. 2012. *Around the Tree: Semantic and Metaphysical Issues Concerning Branching and the Open Future*, volume 361. Springer.

Davidson, Donald. 1971. “Agency”. In Robert Binkley, Richard Bronaugh and Ausonio Marras (eds.), *Agent, Action, and Reason*. Toronto: University of Toronto Press.

Davidson, Donald. 1978. Intending. In Y. Yovel (ed.), *Philosophy of History and Action*. Dordrecht: D. Reidel. 41–60.

Davidson, Donald. 2001. *Essays on Actions and Events*. Oxford: Clarendon Press, 2nd edition.

Donnellan, Keith Sedgwick. 1963. “Knowing What I Am Doing”. *The Journal of Philosophy* 60: 401–409.

Dretske, Fred. 1971. “Conclusive reasons”. *Australasian Journal of Philosophy*, 49(1): 1–22.

Dretske, Fred. 1981. *Knowledge and the Flow of Information*. Cambridge: MIT Press.

Egan, Andy. 2007. “Epistemic Modals, Relativism and Assertion. *Philosophical Studies*, 133(1): 1–22.

Egan, Andy, Hawthorne John and Weatherson Brian. 2005. “Epistemic Modals in Context”. In Gerhard Preyer and Georg Peter (eds.), *Contextualism in Philosophy*. Oxford: Oxford University Press. 131–168.

Falvey, Kevin. 2000. “Knowledge in Intention”. *Philosophical Studies* 99(1): 21–44.

Goldman, Alvin I. 1967. “A Causal Theory of Knowing”. *The Journal of Philosophy*, 64(12): 357–372.

Grice, Herbert Paul. 1972. *Intention and Uncertainty*. Oxford: Oxford University Press.

Haddock, Adrian. 2011. “The Knowledge That a Man Has of His Intentional Actions”. In Anton Ford, Jennifer Hornsby and Frederic Stoutland (eds.), *Essays on Anscombe’s Intention*. Cambridge (MA): Harvard University Press.

Hawthorne, John. 2007. “Eavesdroppers and Epistemic Modals”. *Philosophical Issues*, 17(1): 92–101.

Kaplan, David. 1979. “On the Logic of Demonstratives”. *Journal of Philosophical Logic*, 8(1): 81–98.

Kaplan, David. 1989. “Afterthoughts”. In Joseph Almog, John Perry and Howard Wettstein (eds.), *Themes from Kaplan* (565–614). Oxford: Oxford University Press.

Kim, Kihyeon. 1993. “Internalism and Externalism in Epistemology”. *American Philosophical Quarterly*. 303–316.

Kneer, Markus. 2015. *Perspective in language*. Paris: PhD Dissertation, EHESS ENS Paris.

Kneer, Markus. 2020. Predicates of personal taste, semantic incompleteness, and necessitarianism. *Linguistics and Philosophy* (online first). https://doi.org/10.1007/s10988-020-09303-w.

Kneer, Markus. (in prep). "Knowledge in action: Anscombe v. Davidson."

Langton, Rae. 2004. "Intention as Faith". In John Hyman and Helen Steward (eds.), *Agency and Action*. Cambridge: Cambridge University Press. 243 – 258.

Lehrer, Keith. 2000. *Theory of knowledge*. Boulder: Westview Press.

MacFarlane, John. 2003. "Future Contingents and Relative Truth". *The Philosophical Quarterly*, 53(212): 321 – 336.

MacFarlane, John. 2005. "The Assessment Sensitivity of Knowledge Attributions". *Oxford Studies in Epistemology* 1: 197 – 233.

MacFarlane, John. 2008. "Truth in the Garden of Forking Paths". In M. García-Carpintero and Max Kölbe (eds.), *Relative Truth*. Oxford: Oxford University Press. 81 – 102.

MacFarlane, John. 2011. "Epistemic Modals are Assessment-sensitive". In Andy Egan and Brian Weatherson (Eds.), *Epistemic Modality*. Oxford: Oxford University Press. 1 – 44.

MacFarlane, John. 2014. *Assessment Sensitivity: Relative Truth and its Applications*. Oxford: Oxford University Press.

McDowell, John. 2011. "Anscombe on Bodily Self-knowledge". In Anton Ford, Jennifer. Hornsby and Frederick Stoutland (Eds.), *Essays on Anscombe's Intention*. Princeton: Princeton University Press. 128 – 146.

Mele, Alfred R. and Cushman Fiery. 2007. "Intentional Action, Folk Judgments, and Stories: Sorting Things Out". *Midwest Studies in Philosophy* 31(1): 184 – 201.

Moran, Richard. 2004. "Anscombe on 'Practical Knowledge'". In John Hyman and Helen Steward (eds.), *Agency and Action*. New York: Cambridge University Press. 43 – 68.

Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge: Harvard University Press.

O'Shaughnessy, Brian. 2003. "The Epistemology of Physical Action". In Johannes Roessler and Naomi Eilan (eds.), *Agency and Self-Awareness*. New York: Oxford University Press. 345 – 57.

Paul, Sarah, K. 2009. "How We Know What We're Doing". *Philosophers'Imprint*, 9(11): 1 – 24.

Pickard, Hanna. 2004. "Knowledge of Action Without Observation". *Proceedings of the Aristotelian Society* 104(3): 205 – 230.

Schwenkler, John. 2011. "Perception and Practical Knowledge". *Philosophical Explorations* 14(2): 137 – 152.

Schwenkler, John. 2012. "Non-observational Knowledge of Action". *Philosophy Compass* 7(10): 731 – 740.

Setiya, Kieran. 2007. *Reasons Without Rationalism*. Princeton: Princeton University Press.

Setiya, Kieran. 2008. "Practical Knowledge". *Ethics*, 118(3): 388 – 409.

Setiya, Kieran. 2009. "Practical Knowledge Revisited". *Ethics* 120(1): 128 – 137.

Stephenson, Tamina. 2007. "Judge Dependence, Epistemic Modals, and Predicates of Personal Taste". *Linguistics and Philosophy* 30(4): 487 – 525.

Thomason, Richmond H. 1970. "Indeterminist Time and Truth-value Gaps". *Theoria*, 36(3): 264 – 281.

Thompson, Michael. 2008. *Life and action*. Cambridge (MA): Harvard University Press.

Thompson, Michael. 2011. "Anscombe's Intention and Practical Knowledge". In Anton Ford, Jennifer Hornsby and Frederick Stoutland (Eds.), *Essays on Anscombe's Intention*. Cambridge: Harvard University Press. 198 – 210.

Velleman, David, J. 1989. *Practical Reflection*. Princeton: Princeton University Press.
Von Fintel, Kai and Gillies Anthony S. 2007. "An Opinionated Guide to Epistemic Modality". *Oxford studies in epistemology*, 2: 32–62.
Von Fintel, Kai and Gillies Anthony S. 2008. "Cia Leaks". *Philosophical Review*, 117(1): 77–98.
Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.
Yalcin, Seth. 2011. "Nonfactualism About Epistemic Modality". In Andy Egan and Brian Weatherson (Eds.), *Epistemic Modality*. Oxford: Oxford University Press. 295–332.

Sean Crawford

De Re Explanation of Action in Context, the Problem of 'Near-Contraries' and Belief Fragmentation

Abstract: Commonsense psychological explanation of action upon objects seems to require not only reference to agents' demonstrative beliefs about the objects acted upon but also the *de re* ascription of these demonstrative beliefs. There is an influential objection, however, to the *de re* component: since *de re* ascriptions permit the attribution to agents of inconsistent attitudes about the objects acted upon, they cannot explain (or predict) agents' actions upon those objects. This paper answers the objection by presenting a contextualist theory of *de re* action explanation according to which agents' beliefs about objects are logically fragmented.

1 *De Re* Explanation

Ralph is a contestant on a television game show (somewhat like Monty Hall's *Let's Make a Deal*) in which he is asked to open the door that hides the prize. There are two doors: a red one and a green one. The green one contains the prize (a car, say) and the red one nothing (or maybe a goat). Unlucky Ralph opens the red door. Suppose we want to explain why Ralph did what he did, opened the door that he did. If we want to explain the occurrence of the event that is Ralph's opening the door he did in fact open, that is to say, Ralph's actual door-opening behavior on that occasion, we must choose to describe the event in some way. Observing the scene, we may notice many things about the event we want to explain: Ralph opened *the red door*, Ralph opened *the door on his left*, he opened *the smallest door*, *the door with the goat behind it*, and so on. We must choose one of the highlighted definite descriptions as the singular term we shall use to pick out the object Ralph acted upon in our description of the event to be explained. Suppose we opt to use 'the red door' in our *explanadum* statement. A satisfactory explanation will, in turn, obviously need to employ 'the red door' in the *explanans* statement(s), as opposed to 'the door on his left' or 'the smallest door'. For it is no explanation for Ralph's opening the red door to say he believed the door on his left to contain the prize. So far, so trivial.

Sean Crawford, The University of Manchester (Sean.Crawford@manchester.ac.uk)

https://doi.org/10.1515/9783110702286-010

There is, however, a far from trivial further point to make: there is no way to explain (or predict) Ralph's opening the door in question without attributing to him a belief that relates him to the very door he opened. That is to say, a necessary condition for explaining why Ralph did what he did, that is, why he opened the red door—if that is the description we are using to describe his action—is the citation of the *de re* ascription that the red door was believed by Ralph to harbour the prize (Tomkow 1992; Crawford 2012). (Similarly, a necessary part of the explanation for Oedipus's marrying his mother must be that his mother was believed by him not to be his mother.) We cannot get by with attributing to Ralph a belief *de dicto*, such as that Ralph believed that the red door harboured the prize, where 'the red door' occurs obliquely and therefore is not open to the usual extensional operations. For that belief does not explain Ralph's opening the red door unless the red door is believed by Ralph to be the red door. If Ralph is color blind, then it might be the green door that is believed by him to be the red door, and hence, to hide the prize. In that case, although Ralph thinks to himself *the red door hides the prize*, he thinks this *of* the green door, and so he will actually open the green door—and win the prize.

Though examples of mistakes and unsuccessful actions owing to false beliefs make the need for *de re* ascriptions especially vivid, it is, *au fond*, the *de re* ascription, whether explicit or implicit, that carries the load in *any* explanation of what an agent did or prediction of what an agent will do, whether unsuccessful or successful. Indeed, in cases where an agent does succeed in doing what he is trying to do, it will be a *de re* ascription that explains why he succeeded (Brandom 1994: 523–24). It is the fact that the red door is indeed the door that is believed by Ralph to be the red door that explains why he succeeded in opening it.

2 *De Re E*xplanation Elaborated

In an important though unduly neglected paper, Adam Morton has pointed out that "If we explain actions by appealing to general principles, then we must quantify into psychological contexts to state these principles" (Morton 1975:, 7).[1] His example involves a character Donald explaining his being physically attacked by Leo, by saying that Leo thought he (Donald) had insulted Leo and that

1 A similar claim is made by Peacocke (1981) though somewhat below the surface. The idea was first brought to my attention by Terry Tomkow in graduate seminars at Dalhousie University and in unpublished work (Tomkow 1992) to which I am much indebted.

Leo attacks anyone whom he believes to have insulted him. The 'psychological principle' at work here, which involves quantifying in, appears to be:

(PQI) $\forall x$(Leo believes that x has insulted him → Leo attacks x)

This principle about Leo, combined with the *de re* ascription that it is Donald whom Leo believes to have insulted him, explains Leo's attacking Donald (in the manner of fitting a Hempelian covering-law model of explanation). A couple of comments about (PQI) and its kin are in order.

First of all, for well-known reasons emphasized by Perry (1993), following the work of Castañeda (1999), it is necessary that the expression referring to Leo in the *de dicto* content clause of the antecedent of the conditional (i.e., 'him') be taken as indicating the way Leo thinks of himself. For the force of the generalization is lost if we merely take Leo to believe x to have insulted, say, *Leo*—for Leo may not realize that his name is 'Leo' and thus that it is he who has been insulted; indeed, due to severe amnesia and sensory deprivation he may not know any uniquely distinguishing features of himself expressible without the use of 'I' or 'me'. Presumably, on an occasion of being insulted, Leo thinks to himself "He has insulted *me*" or "*I* have been insulted by him" and it is these so-called 'indexical thoughts' that are essential to his becoming offended and to his attacking his insulter. I shall take this to be implicit in what follows.

Second, Morton describes the principle underwriting the explanation of Leo's behavior as one that involves quantifying *into* a psychological context and it is of course controversial whether such quantification is legitimate. We can avoid controversy here by making use of Quine's (1956) classic proposal about how to make logical sense of quantifications into psychological contexts that resist the substitutivity of identity by paraphrasing the (allegedly) troublesome constructions—e.g., 'Leo believes that Donald has insulted him'—into ones in which the subject of the psychological verb phrase following 'that' is taken out of the scope of the opacity-inducing operator leaving the predicate behind to remain within its scope—e.g., 'Donald is believed by Leo to have insulted him'. Such *de re* paraphrases are straightforwardly quantifiable—'$\exists$x(x is believed by Leo to have insulted him)'—*and* permit substitution of co-referring expressions, precisely because the replacement of the singular term ('Donald') by a variable bound by a quantifier no longer involves quantifying into the scope of 'believes'.[2]

2 For further discussion of Quine's account, according to which quantification and substitution go hand-in-hand, see Crawford (2008) esp. p 78n5.

We can, then, reformulate (PQI) along the following lines:

(P1) $\forall x$(x is believed by Leo to have insulted him $\rightarrow$ x is attacked by Leo).

in which there is no quantification into the scope of 'believes'. The moral to be drawn is not, then, exactly as Morton describes it in the above quotation. Rather, it is that if we appeal to general psychological principles in our psychological explanations, then these principles will involve quantifying over the material objects that the psychological attitude verbs (such as 'believes') and action verbs (such as 'attacks') are directed toward. That is, in the psychological explanations the universally quantified psychological principles will be instantiated by ordinary spatio-temporal particulars, such as people (Donald, for example). In short, then, the commonsense psychological explanation of a subject's action upon an object o often appears to invoke a universally quantified conditional that is instantiated by o. This nicely fits the covering-law model of explanation: the psychological principle at work is the law under which the phenomenon to be explained is subsumed and the *de re* belief ascription that is the instantiation of the psychological principle is the statement of initial conditions. From the psychological principle together with the *de re* ascription we can deduce that the action to be explained took place (more exact details on this below).

Morton rightly points out, however, that "The 'anyone' in [P1] cannot mean that all instances of 'If Leo believes that ... has insulted Leo, then Leo attacks ...' are true. For Leo believes that people have insulted him who are too remote in time and place for him to attack" (1975: 6). Nor can it mean that all instances of the foregoing conditional schema are true in which the blanks are replaced by a distinguished term '*D*' of the Kaplanian variety (see Kaplan [1968]) that ensures that Leo has a belief *about* the object in question, for if "*D* approaches Leo disguised as a dancing bear [we] don't want to predict that Leo will punch *D* on his furry nose" (Morton 1975: 7). As stated, then, (P1) is much too simple to be even approximately true. If Leo believes o to have insulted him, he will not punch o even if o is within striking range unless he also believes o to be within striking range and puts the two beliefs together. That is, for o to get punched by Leo, o must be such as to be believed by Leo to be an insulter of him *and* to be within punching range. Suppose Leo sees Donald make a rude gesture at him, and comes to believe Donald to have insulted him, but that Leo is unable to catch him on the occasion of the insult. The next day Donald is at a costume party dressed up as a dancing bear and is standing directly in front of Leo talking to him. Leo certainly believes, of the person who is dressed up as the dancing bear and talking to him, that is, of Donald, that he is within punching range. (Perhaps Leo is looking for trouble and wishes that the dancing bear would in-

sult him so that he could have an excuse for punching him.) So, at the very moment when Donald is talking to Leo, Donald is believed by Leo to have insulted him and is believed by Leo to be within punching range—yet Leo does not punch Donald. Why not? Part of the reason is that at that very moment Donald is not believed by Leo to have insulted him *and* to be within punching range; indeed, at that moment *no one* is believed by Leo to be an insulter within punching range.[3]

As a first step toward achieving a psychological generalization that is even approximately true we need to see the importance of Leo's knowingly ascribing the properties of having insulted him and being within punching range to the same person. In other words, for Donald to get punched by Leo, Leo at least needs to come to believe Donald to have the complex property of *being an insulter within punching range*—though, as we shall shortly see, this is still insufficient to get Leo to punch Donald. There are, of course, many ways that Leo can come to put these two beliefs together or, what amounts to the same thing, many ways that Donald can come to be believed by Leo to possess the property of being an insulter within punching range. For example, Leo might see Donald getting out of his bear costume and recognize him; or someone might tell Leo that the dancing bear talking to him is the guy who insulted him yesterday. However Leo ends up putting his two beliefs together—or, what amounts to the same thing, however he comes to attribute to Donald the property of being an insulter within punching range—this much is needed to get Leo to punch Donald.

It is important to note at this stage that the possibility that Leo can believe, of Donald, that he is an insulter, and believe, of him, that he is within punching range and yet not believe, of him—or indeed of anyone—that he is an insulter within punching range depends on the fact, first noted by Quine (1956), that

(1) Donald is believed by Leo to have insulted him

and

(2) Donald is believed by Leo to be within punching range

do not imply:

(3) Donald is believed by Leo to have insulted him and to be within punching range.

3 I am indebted here to Tomkow (1992), who was the first to point out, contrary to popular opinion (see note 11 below) just how useful and indeed essential the presence and absence of conjunctive *de re* ascriptions are to the explanation of action and non-action.

Recall Quine's famous story about Ralph and Ortcutt. Even though:

(4) Ortcutt is believed by Ralph to be a spy

and

(5) Ortcutt is believed by Ralph not to be a spy

are both true,

(6) Ortcutt is believed by Ralph to be a spy and not to be a spy.

is false. There is a big difference between (6) and

(7) Ortcutt is believed by Ralph to be a spy and is believed by Ralph not to be aspy.

(3) and (4) imply (7) but not (6).[4]

We need not, in any case, rely on examples involving inconsistent beliefs. For it is clear in Ralph's case that although

(8) Ortcutt is believed by Ralph to be the man at the beach

and

(9) Ortcutt is believed by Ralph to be the man in the brown hat are both true,

it is not true that

(10) Ortcutt is believed by Ralph to be the man at the beach and the man in the brown hat.

Imagine that someone tells Ralph that the man he saw at the beach is the man he saw in the brown hat. As Quine tells the story, this should be shocking news to Ralph. But if (10) is *already* true because (8) and (9) are true, then it cannot be news at all to Ralph. Since the details of Quine's story license (8) and (9) and the subjunctive that if Ralph were to become apprised of the fact that the sentence

4 This is discussed and defended in much greater detail in Crawford (2008).

(which he understands) 'The man is the brown hat is the man at the beach' is true, then he would learn something new, (10) cannot be true.[5]

Returning to Leo and Donald, it is obvious that (3) is not sufficient to get Leo to punch Donald. Suppose Arthur tells Leo that the guy who insulted him is within punching range but teasing him does not tell him which of the many people within punching range is the guy who insulted him (the party is very crowded); he does not, for example, tell him that it is the dancing bear. In this situation (3) is true but since Leo does not know *exactly where* the insulter within punching range *is*, Donald does not get punched. If, however, Arthur does tell him that it is the dancing bear within punching range who insulted him, then is the advent of:

(11) Donald is believed by Leo to be an insulter who is also a dancing bear within punching range.

sufficient to lead to Leo's punching Donald? Only given that there is exactly one dancing bear within punching range, of course. If there is more than one dancing bear within punching range, then Arthur will need to tell Leo exactly which one is the insulter: perhaps it is the tallest dancing bear or the one to Leo's right. If Arthur refuses to tell Leo which bear has insulted him, Leo will have to figure it out for himself, perhaps by launching a search and removing the bear costumes one-by-one in order to find his insulter. Which *de re* ascription is the precise one needed to explain or predict Leo's punching Donald obviously depends on the details of Leo's situation *vis-à-vis* Donald, in particular, whether Donald is recognisable by Leo to be his insulter. The general point is simply that the relevant *de re* ascription will be one that mentions some characteristic of Donald that enables Leo to punch him. As Morton notes, "The relevant characteristic will usually concern the spatial location of the insulter-victim, but need not: any characteristic which in Leo's possession will allow him to attack the person having it will do" (1975: 7–8).

In light of this, we might consider improving upon (P1) by revising it to:

(P2) $\forall x \forall t$(At t, x is believed by Leo to have insulted him* and to be someone who can be attacked there* and then* $\rightarrow$ x is attacked by Leo at t).

in which the asterisks are inserted here as reminders that the expressions they are attached to represent Leo's egocentric concepts of *himself* and the spatio-tem-

5 Cf. Dummett 1978, 1975 (appendix). The role of sentence acceptance and (dis)quotation principles in Quine's theory of belief is discussed at length in Crawford (2008).

poral location of x *with respect to himself* (I shall suppress them in what follows). Indeed, to punch someone successfully one needs pretty precise information about the egocentric location and bodily position of one's victim and when exactly it would be a good time to deliver the blow—just ask Joe Louis. As soon as Leo realizes that the dancing bear is the insulter, he will be preparing himself for the punch by deliberating on the delivery: "Okay, I know it's that damn bear ... if he moves a little to my left and puts his paws down a bit I can get a good right hook in ... there he goes to my left ... his paws are down so I can hit him ... now!"—and down goes Donald.

Well—not quite. For as Evans (1982: 132) has made clear, in many cases "in the absence of an object to anchor our dispositions, we can only make rather gross discriminations of areas or regions in egocentric space." Indeed, surely it was the fact that Louis's dispositions were anchored on Schmeling himself that enabled him to knock Schmeling down three times. In other words, in order for Leo to hit Donald he needs to think of him demonstratively; he needs to think of him as *that man* or perhaps as *that bear* (as indeed I already portrayed him as thinking in the recent soliloquy). It is when Leo thinks, of Donald, as *that bear*, that he is an insulter within punching range that Leo hits Donald.

The idea that reference to a demonstrative mode of presentation of an object plays an essential role in any explanation of action upon that object has been argued for most persuasively by Peacocke (1981) under the title of the 'Indispensability Thesis':

> (IT) No set of attitudes gives a satisfactory psychological explanation of a person's acting on a given object unless the content of those attitudes includes a demonstrative mode of presentation of that object (205–206).

Peacocke goes on to give a detailed account of what a demonstrative mode of presentation (DMOP) is, the thrust of which is that *token* DMOPs are object-dependent: their existence and identity depends on the existence and identity of the objects they present, in the sense that if there is no object then there is no token DMOP and if there is a different object then there is a different token DMOP.[6] Peacocke allows that there can be different tokens of the same *type* of DMOP; indeed, he explains what a token DMOP is by saying that it is what ob-

6 The account is developed further in *Sense and Content* (1983).

tains when a type DMOP is 'indexed' to an object.[7] Since it is token DMOPs that are the constituents of an individual's thoughts, it is they that are relevant to considerations of the psychological explanation of an individual's action, and therefore will be the focus of attention.

In commenting on the psychological explanation for why a certain person grasped a certain container, Peacocke says that "Concerning the agent and the container, the *explanandum* is the fact that the former grasped the latter. The *explanandum* is not a propositional attitude, and indeed if we say the *explanandum* is that *a* grasped *b*, then we must note that the places occupied by '*a*' and '*b*' are transparent. The phrase 'acting on an object' in the Indispensability Thesis is shorthand for any such relational *explanandum* sentence of the form '*Rab* at time *t*'" (Peacocke 1981: 206).

So the *explananda* that fall under the purview of (IT) are the same as those that concern *de re* explanation as adumbrated at the outset (§ 1): namely, the things agents actually end up doing (whether they wanted to or not), which we can call *doings* for short.[8] With this in mind, consider Peacocke's example of the person—we can dub her Roxanne—who wants to go on living and thinks that she will die unless she immediately consumes the contents of the container within reach in front of her and so does consume them. According to Peacocke, the relational fact that constitutes the *explanandum* can be described by the relational sentence 'Roxanne grasped the container' in which the proper name and the definite description occur in transparent position. According to (IT), the *explanans* must make reference to the fact that Roxanne had a thought to the effect that '*that container* contains the pill I need to live'; or, if she thinks a thought of the form ⌜the φ contains the pills I need to live⌝ for any φ that denotes the container, then she must also think a thought of the form ⌜the φ is *that container*⌝—both thoughts in which 'that container' picks out an object presented to the subject in a particular way in perception' (Peacocke 1981: 207; cf. Evans 1982). For her thinking a thought of the form ⌜the φ contains the pills I need⌝ can only explain her acting on the container in front of her is she also thinks ⌜the φ is *that container*⌝.

Since the *explanandum* statement is transparent we can use the singular term 'the container left by the doctor' to describe the object acted on by Roxanne.

---

7 So, unlike the object-dependent *de re* senses of McDowell (1984) and the object-dependent *Ideas* of Evans (1982), to which the notion of type does not apply except in so far as it is a collection of object-dependent tokens, types of DMOPs are not object-dependent.

8 Doings in this sense contrast with *tryings*, which are the things agents are trying to do. For further discussion of the difference between tryings and doings and their respective explanations, see Brandom (1994), ch. 8 and Crawford (2012).

The *explanandum* is then just as well described as 'Roxanne grasped the container left by the doctor'. Clearly, it is not an adequate explanation of this to say that Roxanne thought that that container contains the pills I need to live (together with a statement about Roxanne's conative attitudes). As Peacocke himself points out in a later paper, "Quite generally, explanation of a truth by a given set of states is not preserved by substitution of coextensive predicates in that truth, not even if we add a statement of the coextensiveness to the original explanation" (1993: 208). What I wish to add to this is that an explanation of a truth by a given set of states *is* preserved by substitution of coextensive predicates in that truth so long as the same substitution of coextensive predicates is made in the ascription of the given *explanans* set of states. So, for example, if the transparent *explanandum* statement is 'Roxanne grasps that container', and the key *explanans* statement is 'Roxanne believes that that container contains the pills she needs to live', then, if we substitute 'the container left by the doctor' for 'that container' in the *explanandum* statement, the explanation is preserved so long as we make the same substitution in the *explanans* statement. To be able to do this, however, both *explanandum* and *explanans* statements must be transparent, in the sense that the singular term referring to the object acted upon must occur in a position that is open to substitution of co-referring singular terms.

On Peacocke's account of the logical form of an ascription of a thought containing a token DMOP, the position occupied by the singular term referring to the object that the DMOP is 'indexed' to is "transparent and quantifiable" (1981: 190). If this is so, then (IT) allows for a substitution in the *explanans* statement that corresponds to any substitution made in the *explanandum* statement. So, the explanation of the doing that is described as Roxanne's grasping the container left by the doctor, made fully explicit, is: Roxanne believed, of the container left by the doctor, as *that container* (or: under the mode of presentation *that container*), that it contained the pills she needs to live. Though Peacocke's (IT) clearly allows for the 'set of attitudes' in which reference is made to a DMOP to be ascribed relationally, because of the transparency of the 'indexing' position, and therefore to explain doings, he does not highlight the all-important fact that the *relationality* too is absolutely essential to the explanation. The emphasis falls on the fact that reference to DMOPs is needed, which is of course correct. But if the thesis of the explanatory necessity of *de re* (or relational) ascriptions is also correct (as set out in § 1), then it is essential that these DMOPs be ascribed relationally in a *de re* ascription.

Returning to Leo, I propose that our quest for the generalization that is needed to explain his attacking Donald can end with this following psychological generalization:

(P3) ∀x∀t[Bel(Leo, <x>, t, ⌜Insulter-within-attacking-range[y]Person(y)⌝)
→ Attacks(Leo, <x>, t)],

which is rendered in Burge's (1974, 1977) notation for demonstrative constructions, where 'Bel' stands for 'believes'. (P3) can be read as saying that: For any object x and any time t, if, at t, x is believed by Leo to be an insulter within attacking range, and Leo thinks of x at t as *that person*, then x is attacked by Leo at t. The open sentence in the fourth argument place of the antecedent, ⌜Insulter-within-attacking-range[y]Person(y)⌝, which can be rendered in the vernacular as *that person is an insulter within attacking range*, makes reference to the fact that Leo has a perceptual demonstrative mode of presentation of x on the basis of which he is able to locate x.[9] The 'relevant characteristics,' as Morton calls them, of Donald that allow Leo to attack him are precisely those characteristics that are presented to Leo by a (visual) demonstrative mode of presentation of him.

(P3) is a canonical representation of a philosophical precisification of a commonsense principle that is akin to a lawlike causal generalization. (P3) combines with a *de re* belief ascription which is its instantiation to explain a doing by subsuming it in the manner of a covering law explanation, in which the *de re* or relational belief ascription is the statement of initial conditions, to give us something like:

(P3) ∀x∀t[Bel(Leo, <x>, t, ⌜Insulter-within-attacking-range[y]Person(y)⌝)
→ Attacks(Leo, <x>, t)],
(I) Bel(Leo, <Donald>, t_1, ⌜Insulter-within-attacking-range[y]Person(y)⌝)
(E) Attacks(Leo, <Donald>, t_1).

This, then, is a paradigm of what a *de re* explanation, made fully explicit, looks like. The two essential factors in (I) that enable the explanation to work are: (i)

9 The first 'y' in square brackets is not an operator binding the second 'y', which is free and represents the demonstrative 'that' (or other pronoun). The first 'y' marks the scope of the demonstrative, indicating in this case that the property of being an insulter within attacking range is being predicated of *that person* rather than just of *that—i.e.*, the subject of the predication is being singled out by a complex demonstrative (demonstrative plus noun phrase), so 'person' is not occurring purely predicatively (unlike 'being an insulter within attacking range', which is) but is also serving to secure (or guide) reference for (purely predicative) predication. Obviously, exactly which noun phrase figures in the complex demonstrative expression will depend on the circumstances and the explanatory interests and background knowledge of the ascriber. Instead of 'that person' it might be 'that dancing bear'. The important point is that there is always a crucial constant element, the demonstrative, simple or complex. For further explanation and discussion of this aspect of the Burgean view, including the notation, see, aside from Burge's own work already cited, the beginning of § 2 of Crawford 2018, esp. note 9.

the relationality of the *de re* ascription, in which the singular term referring to the object acted upon is in referential position outside the psychological verb, and (ii) the reference to the subject's perceptual demonstrative mode of presentation of the object acted upon.[10]

3 The Case Against *De Re* Explanation

My conclusion so far is that if we are to explain or predict what an agent will do, that is, which object an agent will act on, we shall need some *de re* ascriptions of thought that relate the agent to the object acted upon and which include a reference to a perceptual demonstrative mode of presentation of the object acted upon. Both features—the relationality of the *de re* ascription and the inclusion within it of a demonstrative expression—are necessary. Moreover, it appears that in our commonsense explanations of people's doings we regularly appeal to causal generalizations that quantify over the objects of their doings. For these generalizations to play the explanatory role they appear to play their antecedents must be instantiated by *de re* thought ascriptions. These relational thought ascriptions combine with the psychological generalizations, in the manner of a covering-law explanation, to explain the event of an agent's doing. Whether one thinks that a covering-law (or generalization) is not always or even ever required in the explanation of a doing, or whether one thinks that such alleged cases are merely elliptical and tacitly presuppose a covering law, the fact remains that there does not appear to be any way of explaining a doing without at least citing some *de re* ascriptions. While there may be situations in which principles like (P3) are otiose, relational psychological statements like (I) clearly are not.

Yet the very viability of *de re* explanation has been questioned: specifically, the relational component—the use of *de re* ascriptions—has been widely denigrat-

10 As Hempel (1988) pointed out in one of his late papers, for such explanations to be deductive the *explanans* must include a statement, which Hempel calls a 'proviso,' to the effect that no factors not mentioned in (P3) that are relevant to the outcome of the event described by (E) are present, i.e., that (L) states the whole truth about the relevant circumstances present. Consideration of the issue of provisos and '*ceteris paribus*' or 'hedged' laws is beyond the scope of this paper. See Cartwright 1983, Hempel 1988, Fodor 1989, and Pietrosky & Rey 1995. There is further debate about whether the provisos needed in psychological explanation differ in kind or only in degree from those found in other explanatory schemes—on this see the debate between Fodor (1987, chapter 1; 1989) and Davidson (1980, 1987, 1993)—but I must leave the issue moot here.

ed by philosophers.[11] The argument against the explanatoriness of *de re* ascriptions appears to turn somehow on the fact that they, unlike *de dicto* ascriptions, are not governed by principles of logic and rationality (Crawford 2012). Dennett (1982: 199 – 200), for example, correctly notes something we have already seen, namely, that from the fact that *a* is believed by him to be *F* and *a* is believed by him to be *G* it does not follow that *a* is believed by him to be *F* and *G*. As we saw above, however, it is precisely this fact that we exploited to our advantage in giving a *de re* explanation of why Leo did not punch Donald when he was dressed up as a dancing bear, that is, why Leo did not punch Donald even though he believed him to have insulted him and believed him to be within punching range. So the fact that the predicative content clauses of *de re* ascriptions do not abide by conjunction introduction is not a problem for *de re* explanation. It is rather their recalcitrance with respect to the other logical rules of consistency that Dennett mentions that poses the threat to *de re* explanation. There are two hard cases to confront. The first is the one originally noted by Quine in which someone believes of *a*, that it is *F* and believes of *a* that it is not *F*. The second case Dennett mentions is one in which an individual thinks, of one thing *a*, that it is *the only F*, and thinks of another thing *b* (where $a \neq b$), that it is *the only F*. Another type of case in all relevant respects similar to this second type of case is where the individual thinks of *a*, that it is *the more F* or *the most F*, and thinks of *b* (where $a \neq b$), that it is *the more F* or *the most F*, where *F* is an adjective such as 'beautiful' or 'valuable'. For our purposes, the important points to notice are the following. The first type of case involves a *single* object to which a person unknowingly ascribes *two* inconsistent properties. The second case involves *two* (or more) objects to each of which a person unknowingly ascribes, inconsistently, a *single* uniquely instantiated property. For convenience, I shall refer to both types of cases as cases of 'near-contraries,' meaning by this cases of inconsistent attitudes that a rational person may have. Though one can generate cases of near-contraries with attitudes other than belief—e. g., a person may intend, of *a*, that he act on it and intend, of *a*, that he not act on it (think of Oedipus *vis-à-vis* his mother and father)—I shall stick with belief.

Constructing out of these materials the precise argument against *de re* explanation that its critics have in mind is not a straightforward matter; indeed, the argument, whatever exactly it is supposed to be, is rarely made explicit. It

11 E. g., by Schiffer 1978, Fodor 1980, Dennett 1982, Baker 1982, Lycan 1985, Grandy 1986, Carruthers 1988, and Boghossian 1994. Cf. Burge 1982. Baker 1982 gives by far the most impressive argument against *de re* explanation and so it is her views I shall focus on. The demonstrative component is not in question here, so I shall ignore it in what follows.

often seems to proceed along the following rather vague lines.[12] *De re* ascriptions, suffering as they do from the logical looseness just remarked on, do not abide by principles of logical consistency. Unlike the *de dicto* ascription of beliefs, the *de re* ascription of beliefs is not done on the basis of the so-called 'assumption of rationality.' On the *de dicto* way of ascribing beliefs, it is (by and large) supposed to follow from the fact that Roxanne believes there are zebras in Africa that she does not also believe that there are no zebras in Africa.[13] Speaking *de re*, however, it does not follow from the fact that zebras are believed by Roxanne to live in Africa that zebras are not also believed by Roxanne not to live in Africa. In short, *de dicto* ascriptions are governed by the assumption of rationality whereas *de re* ascriptions are not. The 'assumption of rationality' is no doubt in need of detailed explication; but for our purposes it will suffice to take it to mean the assumption that the basic logical rules of consistency that are flouted by the two types of near-contraries apply to the endeavor in question. So, the argument against *de re* psychological explanation its critics appear to have in mind is that (1) psychological explanation is rationalization, in the sense that its coherence depends on the assumption of rationality; (2) *de re* ascriptions are not governed by the assumption of rationality; therefore, (3) *de re* ascriptions cannot be used in psychological explanation.

The general response to this general line of argument is as follows. *De re* psychological explanation is not a form of rationalization, so the fact that *de re* ascriptions do not abide by the assumption of rationality is no objection to their explanatory credentials. One could put the reply this way. The *explanandum* of a rationalizing psychological explanation is an action under a description under which it is intentional (i.e., a *trying*); this is the kind of *explanandum* that invokes an agent's reasons, and hence, the *de dicto* attribution of thoughts. The *explanandum* of a *de re* psychological explanation is an action *sans phrase* (i.e., a *doing*); such an *explanandum* does not (wholly) invoke an agent's reasons, and hence, the fact that a *de re* ascription does not specify an agent's rea-

12 Other than Dennett, this line of thought is most explicit in Schiffer 1978 and Boghossian 1994.

13 Schiffer (1978: 138). Although Schiffer speaks of the 'internal functional role' of a belief it is clear that this role is defined by rational or broadly logical relations. Davidson (1980, 1984) is, of course, famous for emphasizing the assumption of rationality (which forms part of the principle of charity) on the basis of which we ascribe thoughts to others. It is clear that he is talking about *de dicto* ascriptions. Ditto Dennett (1987).

sons for acting is no objection to *de re* ascriptions being used in these kinds of explanations.[14]

It is not, however, very satisfying to leave matters here. So we need to work through the details of an example of each type of near-contrary. We need, first, to see exactly how each type appears to cause problems for *de re* explanation and, second, to see whether we have the specific resources to answer these problems. To this end, I turn now to the most sophisticated argument against *de re* explanation to be found in the literature: Lynne Rudder Baker's article '*De Re* Belief in Action.'[15] Baker (1982) argues that *de re* ascriptions cannot figure in action explanations. According to her, "The diagnosis of the difficulty ... reveals that its source is surprisingly deep; indeed, it is so fundamental that ascriptions of *de re* attitudes regarding concrete objects threaten the coherence of any explanation containing them" (366). I shall rebut Baker's charge by showing that *de re* ascriptions do not in fact lead to explanatory incoherence; on the contrary, they are, as we already have reason to believe, essential in the explanation of action.

It is worth noting first something rather paradoxical about the position we find ourselves in: on the one hand we appear to have an argument for the necessity of *de re* explanation, while on the other we appear to have an argument—or rather the semblance of one—against the very possibility of *de re* explanation. Recall that part of the argument for the explanatory necessity of *de re* ascriptions was that *de dicto* ascriptions are insufficient to explain doings. Recall the example of Ralph and the green door that I opened with (in § 1). Baker herself agrees that *de dicto* ascriptions cannot explain action upon an object (Baker 1982: 380). But now, if *de re* ascriptions too are explanatorily impotent, as she also claims, then what kind of ascriptions will do the job? Baker does offer her own theory as to what type of ascriptions are needed. Unfortunately, her theory is a version of the dual component conception of explanation, which is fundamentally flawed (Crawford 2012: 795 – 6). I shall try to show that since *de re* ascriptions are not explanatorily impotent, her extremely complex account is anyway otiose.[16]

14 Cf. note 8 above. This summarizes what is discussed and defended at much greater length in Crawford (2012).

15 Besides its impressive negative brief against *de re* explanation, this article also contains the earliest positive presentation of the dual-component model for explaining action upon objects—anticipating, and, it must be said, considerably outstripping in its sophistication the later dual-component theories of Noonan 1986, 1991, Carruthers 1987, and Segal 1989. I criticize the dual-component model of action explanation in Crawford 2012: 793 – 97.

16 Baker, worrying about her own account, asks: "Is it plausible that ordinary intentional explanation rests on such abstruse principles, which seem to presuppose sophisticated conceptual apparatus on the part of the agent?" (1982: 386). In my opinion, the answer to this question is a resounding no.

In order to present the argument against *de re* explanation, I shall work with a challenging example of Baker's which she adapts from Herbert Heidelberger (1979):[17]

> [Jones] is asked to remove the more valuable of two objects on a table in front of him. On the table are a carved jade dish and a painted porcelain basket, which [Jones] believes are of unequal value. ... After inspecting the dish and the basket, Jones says ... "I believe that this [demonstrating the basket], but not this [demonstrating the dish] is the more valuable". ... Jones then says that he would feel more confident of his judgement if he were allowed to pick up the objects and handle them. To oblige Jones without giving him too great an advantage over Smith [a competitor], he is allowed to handle the objects, but only on the condition that he be blindfolded. Now the two objects are quite similar in shape; and although they differ in texture, the way that they differ in texture is not obvious to the eye. (The painting on the porcelain basket is quite skilful.) Jones, however, sure of his ability to distinguish the dish from the basket by touch alone, consents to the blindfold and gives each object a thorough tactual examination, which, he thinks, confirms his earlier judgement based on the visual examination. So believing that he is holding the basket, Jones says confidently, "I believe that this [demonstrating the object he is holding] is the more valuable." It turns out, of course, that it is the dish that Jones is holding. ... So without changing his mind, Jones comes to believe of the two objects that each is the more valuable. The dish replaced, the blindfold discarded, Jones is now asked to remove the more valuable of the two objects on the table in front of him. ... Jones (predictably) removes the basket (Baker 1982: 367–69).

As we have already seen (§ 1), knowing that Jones intends to choose the more valuable item and that he believes *de dicto* that the more valuable item is the basket does not necessarily allow us to predict that he will choose the basket. For when he is blindfolded he removes the dish, even though the aforementioned *de dicto* ascription is true of him as he is doing so. What appears to tell us what he will do (as opposed to what he will try to do) is the *de re* ascription that it is the dish that is believed by him to be the more valuable; and when he is not blindfolded, that it is the basket that is believed by him to be the more valuable.

But as soon as we dissociate a belief ascription from Jones's conception of the things and train it upon ours, as we do when we make the aforementioned *de re* ascriptions, we run up against the fact that Jones has inconsistent beliefs. For, from our perspective, he not only believes of the dish that it is the more valuable but also believes of the basket that it is the more valuable.[18] Moreover, as

17 This type of example was first presented, so far as I know, by Richard Feldman 1978.

18 For more detailed discussion of how *de re* and *de dicto* ascriptions are linked to the perspectives (or commitments) of the ascriber and ascribee, respectively, see Brandom (1994, ch. 8) and

soon as we shift to the *de re* style of ascription, it appears that we must allow that Jones also believes of the dish that it is *not* the more valuable and believes of the basket that it is *not* the more valuable. All these *de re* ascriptions appear to stand and fall together; if any one is correct, then all are. But this appears to undermine the suggestion that we are able to predict Jones's actions on the basis of *de re* ascriptions. The problem is obvious: since the basket is also believed by him to be the more valuable item, how can his belief, of the dish, that it is the more valuable item, explain his removing the dish rather than the basket, when he is blindfolded—and vice versa when he is not blindfolded? The difficulty here is in fact twofold: for how can Jones's belief of the dish (or basket), that it is the more valuable item explain his removal of the dish (basket) if, given his intention to remove the more valuable item, he also believes of the dish (basket) that it is *not* the more valuable item? It appears that neither *de dicto* nor *de re* ascriptions provide the materials for a prediction or explanation of Jones's behavior. Yet, as commonsense psychologists, we appear able to predict his behavior with relative ease. How do we do it?

Let us zero in on the relevant *de re* ascriptions which Baker takes to be true of Jones:

(12) The basket is believed by Jones to be the more valuable. [via sight]
(13) The dish is believed by Jones not to be the more valuable. [via sight]
(14) The dish is believed by Jones to be the more valuable. [via touch]
(15) The basket is believed by Jones not to be the more valuable. [via touch]

One way of putting the difficulty for the *de re* theorist then is this. First, since (15) is true, citing (12) cannot explain Jones's removing the basket. This obviously corresponds to the first type of case of near-contraries. Second, as Baker puts it, "since Jones has the same belief of the dish, his belief of the basket that it is the more valuable cannot explain his removing the basket rather than the dish" (1982: 369–70). Given Jones's intention to remove the most valuable item, (12) cannot explain his removing the basket because (14) is also true. This is an instance of the second type of case of near-contraries. So, the case against (12)'s explaining Jones's removal of the basket is overdetermined: both (14) and (15) conflict with the claim that (12) explains or predicts Jones's removal of the basket. The fundamental problem is that the practice of attributing attitudes in the *de re* way permits, by definition, the attribution of near-contraries. Baker claims that the problems raised by such examples are "completely gener-

Crawford (2012). Brandom, however, curiously fails to register the existence of near-contrary *de re* ascriptions.

al" and that, "It is not just that we have failed to ascribe the right *de re* belief regarding the object acted upon; for the problems stem from the possibility of near-contraries, which haunts *all de re* beliefs regarding objects" (372).

4 In Defense of D*e Re* Explanation

In spite of the fact that Baker lists only (12)–(15), together with Jones's intention to remove the most valuable item and his ability to do so, she remarks that "Jones (*predictably*) removes the basket" (when not blindfolded; my emphasis); elsewhere she describes Jones's action as "intuitively unpuzzling" (1982: 378). The question then arises: if (12)–(15) are insufficient to predict Jones's removal of the basket then how are we able to do it? How do we know he will remove the basket and not the dish, given that he believes the basket not to be the more valuable and believes the dish to be the more valuable?

Well, as a first step in the right direction, it quite obviously has something to do with the facts that (a) the *de re* ascriptions (12)–(15) become true in two different contexts or circumstances: one in which the items are presented to Jones visually and one in which the items are presented to him tactually; and (b) Jones performs his removal in one of these contexts, that is, on the basis of his *visual* perception of the dish and basket rather than his *tactual* perception of them. If he had been asked to remove the more valuable item while still blindfolded, and hence on the basis of his tactual perception of the respective items, then he would have removed the dish instead—even though he believes the dish not to be the more valuable and believes the basket to be the more valuable. This suggests that the difference between the actual and counterfactual situations, just before Jones removes the respective items, has something essentially to do with the interplay between the context of belief formation, in which the *de re* ascriptions become true, and the context of action in which Jones's removal takes place. Indeed, all the various cases of mistaken identity involving subjects' states of mind before their potential recognition that have paraded through the philosophical literature that appear to confound *de re* explanation: Jones and the basket and dish, Ralph and Ortcutt, Leo and Donald, Oedipus and Jocasta, the Babylonians and Venus, even Oscar and H_2O and XYZ—all these turn on the protagonists not recognizing something as something they believe to have a certain property because of a disparity between the original context in which the *de re* ascriptions true of them become true of them and the new context of their subsequent actions. This indicates that the reason why we know just what these protagonists are going to do must have something essentially to do with our knowledge of the context in

which the *de re* ascriptions became true and the context in which their subsequent actions are taking place and the interplay between these two contexts.

In answer to the question of how Jones's belief, of the basket, that it is the more valuable, can explain his removing it if he also believes, of the basket, that it is not the more valuable, my suggestion is that (12) is *relevant* to an explanation of Jones's removing the basket in a context in which it is visually presented to him because it became true in a context in which the basket was visually presented to Jones. The fact that (15) is also true of Jones—even as he is removing the basket—has no relevance in an explanation of Jones's removing the basket in the context of its being visually presented to him because (15) did not become true in a context in which the basket was visually presented to him. Since (15) became true in a context in which the basket was tactually presented to Jones, it is not relevant to an explanation of his removing the basket in a context in which the basket is not tactually presented to him. The same thing can be said in response to the second type of near-contrariety. Given the interplay and disparity between contexts, we can say that (14) does not conflict with the fact that (12) explains Jones's removing the basket on the basis of his visual perception of the items because (14) became true in virtue of Jones's tactual perception of the items. Since the action of Jones's that we are explaining takes place in a context different from that in which (14) became true there is no reason why (14) should be relevant to this explanation.

We can make more precise the idea that it is the interplay and disparity between contexts of action and contexts of belief formation that renders some *de re* ascriptions explanatorily relevant and others explanatorily irrelevant by drawing on Sosa and Pastin's (1981) concepts of motivating and locating properties.[19] Reverting to Quine's simpler tale of Ralph and Ortcutt for a moment, consider that, speaking *de dicto*, Ralph believes that the suspicious man in the brown hat is a spy. He says as much. Indeed, he says more: he says that the next time he sees the suspicious man in the brown hat he will contact the FBI. So, the property of Ortcutt that *motivates* Ralph to contact the FBI is having-the-suspicious-man-in-the-brown-hat appearance. Generally speaking, a property F motivates an agent S to φ an object x at t if and only if at t Fx and at t S wants to φ the thing with F. This is not enough to actually get Ralph to contact the FBI, however. He also needs to *locate* Ortcutt on the basis of that motivating property. After all, there is Ralph talking to Ortcutt on the beach and telling him that the next time he sees the man in the brown hat he is going to contact the FBI. A property F locates

19 The basic idea, however, is clearly present already in Morton (1975), as should be evident from the earlier discussion of universally quantified psychological principles.

x for S to φ at t if and only if at t Fx and at t S believes he is able to φ the thing with F. Ralph's problem is that he is motivated to inform on Ortcutt on the basis of a property that does not locate Ortcutt for him. When he is talking to Ortcutt on the beach, the property that locates Ortcutt for him is having-the-man-at-the-beach appearance. So there is a property of Ortcutt that motivates Ralph to inform on him and a property of Ortcutt that locates him for him, but the properties are not the same. For Ralph to inform on Ortcutt, there needs to be a *single* property of Ortcutt that both motivates Ralph to inform on him and that locates him for informing upon. This is why, so far in the story, he will inform on him only when he sees him acting suspiciously in the brown hat. Again, generally speaking, we can say that for S to φ x at t there must be a property F of x which both motivates S to φ x at t and locates x for S to φ at t.

Back to Jones. Consider, in particular, the fact that, speaking *de dicto*, Jones thinks that the basket is the more valuable. It is clear that after his visual and tactual inspections, he would assent, if queried, to the sentence 'The basket is the more valuable'. (Recall that when blindfolded and demonstrating the dish as the item he thinks to be the more valuable Jones mistakenly thinks he is demonstrating the basket and thereby confirming his original visual judgement that the basket is the more valuable.) The basket looks like the basket to Jones and the dish looks like the dish to Jones; unfortunately for Jones, however, the dish feels like the basket to him and the basket feels like the dish to him. So, there are only two properties that motivate Jones to remove one of the items: having a baskety appearance or having a baskety feel. There are four properties that locate the items for Jones: having a baskety appearance, having a dishy appearance, having a baskety feel, and having a dishy feel. When Jones is looking at the items he removes the basket because only it is located for him by a property that also motivates him to remove it. When Jones is feeling the items he removes the dish because only it is located for him by a property that also motivates him to remove it. When Jones is looking at the items, the basket is located for him by its baskety appearance and the dish is located for him by its dishy appearance. However, of these two locating properties only one of them is also a motivating property, namely, having a baskety appearance, and so Jones removes the basket. Just the reverse is true when Jones is feeling the items. The basket is located by its dishy feel and the dish is located by its baskety feel. Of these two properties only one is a motivating property: having a baskety feel. Since only the dish has a property that locates it for Jones and that motivates Jones to remove it, Jones removes the dish.

The idea of a motivating property is another way of describing the relevance of the context of belief fixation to action on the basis of belief; the idea of a locating property is another way of describing the relevance of the context of ac-

tion to action on the basis of belief. For an agent to act on an object at a certain time that object must present some property to the agent that both locates the object for him to act on at that time and motivates him to act on it at that time. A property of an object becomes a motivating property in the context of belief formation. This is what has happened with Jones: it is when (12) becomes true that having a baskety appearance becomes a motivating property for Jones. It is when (14) becomes true that having a baskety feel becomes a motivating property for Jones. If either of these motivating properties is to lead Jones to remove an item then they must also be locating properties. It is when one of these motivating properties becomes a locating property that action occurs and Jones removes one of the items. When the context of action is one in which Jones is looking at the items, then of the two locating properties, having a baskety appearance and having dishy appearance, only the former is also a motivating property. So only (12) is relevant to an explanation of Jones's removing the basket on the basis of sight. The same applies *mutatis mutandis* for the context in which Jones is touching the items while blindfolded.

5 *De Re* Fragmentation and Tracking

By way of concluding, I would like briefly to relate this solution to the problem of near-contraries to a theme in the work of Stalnaker (1984) and Lewis (1986). Because they both hold that beliefs are constitutively tied to behavior, neither Stalnaker (who calls his approach to intentionality the 'pragmatic picture') nor Lewis (who is a commonsense functionalist) think it makes sense to speak of individual beliefs that are held singly. In their view, beliefs come in large clumps: Lewis calls them 'belief systems'; Stalnaker refers to them as 'belief states.' Furthermore, Stalnaker claims that "It is compatible with the pragmatic account that the rational dispositions that a person has at one time should arise from several different belief states. A person may be disposed, in one kind of context, or with respect to one kind of action, to behave in ways that are correctly explained by one belief state and at the same time be disposed, in another kind of context, or with respect to another kind of action, to behave in ways that would be explained by a different belief state. This need not be a matter of shifting from one state to another or vacillating between states; the agent might, at the same time, be in two stable belief states, be in two different dispositional states which are displayed in different kinds of situations. ... But if an agent can be in distinct belief states at the same time ... then there is no reason why these belief states cannot be incompatible. In such a case an agent would believe both a proposition and its contradictory, but would not therefore believe everything.

It would still be possible in such a situation to explain the agent's actions as rational actions according to the usual pattern" (1984: 83).

Lewis says much the same thing. In his own words, "Belief is compartmentalized and fragmented" (1986: 30). Of course, *de re* explanation is not, as I have emphasized, a form of rationalization, since *de re* ascriptions do not give the way the believer thinks of things and are not bound by principles of rationality. So agreeing with Stalnaker's basic point that incompatible fragmented beliefs can still explain action does not commit me to his claim that this kind of explanation is the 'usual' pattern of rationalization. Indeed, it is precisely the fact that *de re* explanation is not a form of rationalization (Crawford 2012), and that *de re* ascriptions are not subject to the assumption of rationality, that gives my use of the idea of fragmentation an independent motivation not clearly open to Stalnaker. Whereas he is motivated to hold that beliefs are fragmented in order to save his theory of belief from unpalatable consequences (namely, the problems of deduction and logical omniscience acutely faced by his possible worlds approach), we can say that *de re* belief ascriptions are fragmented (or compartmentalized or partitioned), and the motivation for this is nothing other than the uncontroversial fact, agreed by virtually all parties, that *de re* ascriptions are not governed by the assumption of rationality.

The logical fragmentation of the *de re* ascriptions (12)–(15) helps to explain how (12) can explain Jones's removing the basket when the near-contrary (15) is true. The very fact that it does not follow from (12) and (15), for example, that Jones believes of the basket, that it is the more valuable and not the more valuable, shows that (12) and (15) stand compartmentalized or partitioned from each other, and hence, stand ready to be employed separately in different explanatory circumstances.

One important question that remains, however, is just how finely discriminated these 'contexts' are to which appeal has been made in order to solve the problem of near-contraries. Take Jones and the basket and the dish. I have said that, basically, the reason why Jones does not remove the basket when he is feeling it is that this 'context of action' is different from the 'context of belief formation' in which the *de re* ascription that he believes the basket to be the more valuable becomes true. But on what ground can I claim this? What explains why this is a different context, other than the fact that he does not remove the basket in it—which, of course, is the very thing to be explained? This is a difficult question involving what Kaplan (1989) has called 'cognitive dynamics.' Taking a cue from Evans (1981, 1982), I suggest that the answer lies in the fact that Jones has *lost track* of the basket at some point between the earlier time when he is looking at it and the later time when he is touching it and the still later time when he is looking at it again. This is also what has happened to Ralph *vis-à-*

vis Ortcutt: he loses track of him between the time when he sees him in the brown hat and the time when he sees him at the beach. So one plausible proposal about how to discriminate contexts is by appeal to the notion of keeping track of an object through space and time. A subject moves into a different context with respect to an object when he loses track of that object. So, if Jones had not been blindfolded, and was allowed to touch the items when he was also looking at them, then he would not have lost track of the basket. In this situation, he would have removed the basket both on the basis of sight and on the basis of touch. In such a case, since he has not lost track of the basket, the contexts of action and belief formation are one and the same. We can imagine a similar situation with respect to Ralph: he might have tracked Ortcutt throughout the period in which he changed from his spy-in-a-brown-hat garb into his pillar-at-the-beach garb. In such a situation the context of action and the context of belief formation would be one and the same and so Ralph would have contacted the FBI when he saw Ortcutt at the beach. To put it in terms of the earlier notion of a motivating property, if Ralph had kept track of Ortcutt then another property of Ortcutt would have become a motivating property for Ralph, namely, having-the-pillar-at-the-beach-appearance. In this situation Ralph would have contacted the FBI at the beach and so his context of action and his context of belief formation would not have diverged in the way they did in Quine's original story.

References

Baker, Lynne Rudder. 1982. "*De Re* Belief in Action". *Philosophical Review* 91: 363 – 87.

Boghossian, Paul A. 1994. "The Transparency of Mental Content". *Philosophical Perspectives* 8: 33 – 50.

Brandom, Robert. 1994. *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, (MA): Harvard University Press.

Burge, Tyler. 1974. "Demonstrative Constructions, Reference, and Truth". *Journal of Philosophy* 71: 205 – 23.

Burge, Tyler. 1977. "Belief *De Re*". *Journal of Philosophy* 74: 338 – 62.

Burge, Tyler. 1982. "Other Bodies". In A. Woodfield (ed.) *Thought and Object*. Oxford: Clarendon Press.

Carruthers, Peter. 1987. "Russellian Thoughts" *Mind* 96: 18 – 35.

Carruthers, Peter. 1988. "More Faith than Hope: Russellian Thoughts Attacked". *Analysis* 48: 91 – 96.

Cartwright, Nancy. 1983. "*How the Laws of Physics Lie*". Oxford: Clarendon Press.

Castañeda, Hector-Neri. 1999. *The Phenomeno-Logic of the I.*, James G. Hart and Tomis (eds.), Indiana University Press.

Crawford, Sean. 2008. "Quantifiers and Propositional Attitudes: Quine Revisited" *Synthese* 160: 75 – 96.

Crawford, Sean. 2012. "*De Re* and *De Dicto* Explanation of Action". *Philosophia* 40: 783–798.
Crawford, Sean. 2020. "Perceptual Demonstrative Thought: A Property-Dependent Theory". *Topoi* 39: 439–457.
Davidson, Donald. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.
Davidson, Donald. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
Davidson, Donald. 1987. "Problems in the Explanation of Action". In Philip Pettit, Richard Sylvan and Jean Norman, (eds.), *Metaphysics and Morality: Essays in Honour of J.J.C. Smart*. Oxford: Basil Blackwell.
Davidson, Donald. 1993. "Thinking Causes". In J. Heil and A. Mele (eds.) *Mental Causation*. Oxford: Clarendon Press.
Dennett, Daniel. 1982. "Beyond Belief". In A. Woodfield (ed.) *Though and Object*. Oxford: Clarendon Press.
Dennett, Daniel. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.
Dummett, Michael. 1975. "What is a Theory of Meaning?" In Samuel Guttenplan (ed.) *Mind and Language*. Oxford: Clarendon Press.
Dummett, Michael. 1978. "Frege's Distinction between Sense and Reference". In Michael Dummett, *Truth and Other Enigmas*. London: Duckworth.
Evans, Gareth. 1981. "Understanding Demonstratives". In H. Parret and J. Bouveresse, (eds.) *Meaning and Understanding*. Berlin: de Gruyter.
Evans, Gareth. 1982. *The Varieties of Reference*. Ed. John McDowell. Oxford: Clarendon Press.
Feldman, Richard. 1978. "Actions and *De Re* Beliefs". *Canadian Journal of Philosophy* 8: 577–82.
Fodor, Jerry. 1980. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology". *The Behavioural and Brain Sciences* 3: 63–73.
Fodor, Jerry. 1987. *Psychosemantics*. Cambridge, (MA): MIT Press.
Fodor, Jerry. 1989. "Making Mind Matter More". *Philosophical Topics* 17: 59–79.
Grandy, Richard. 1986. "Some Misconceptions about Belief". In Richard Grandy and Richard Warner (eds.) *Philosophical Grounds of Rationality*. Oxford: Clarendon Press.
Heidelberger, Herbert. 1979. "The Self-Presenting".*Grazer Philosophische Studien* 7: 59–76.
Hempel, Carl G. 1988. "Provisoes: A Problem Concerning the Inferential Function of Scientific Theories". *Erkenntnis* 28: 147–64.
Kaplan, David. 1968. "Quantifying In", *Synthese* 19: 178–214.
Kaplan, David. 1989. "Demonstratives". In J. Almog, J. Perry, and H. Wettstein (eds). *Themes from Kaplan*. New York: Clarendon Press.
Lewis, David. 1986. *On the Plurality of Worlds*. Oxford: Basil Blackwell.
Lycan, William. 1985. "Thoughts about Things". In M. Brand and R.M. Harnish (eds.) *The Representation of Knowledge and Belief*. Tucson: University of Arizona Press.
McDowell, John. 1984. "*De Re* Senses". *Philosophical Quarterly* 34: 283–94.
Morton, Adam. 1975. "Because He Thought He Had Insulted Him". *Journal of Philosophy* 62: 5–15.
Noonan, Harold. 1986. Russellian "Thoughts and Methodological Solipsism". In Jeremy Butterfield (ed.) *Language, Mind, and Logic*. Cambridge: Cambridge University Press
Noonan, Harold. 1991. "Object-Dependent Thoughts and Psychological Redundancy". Analysis 51. 1–9.
Peacocke, Christopher. 1981. "Demonstrative Thought and Psychological Explanation". *Synthese* 49: 187–217.

Peacocke, Christopher. 1983. *Sense and Content*. Oxford: Clarendon Press.
Perry, John. 1993. *The Problem of the Essential Indexical and Other Essays*. Oxford: Oxford University Press.
Pietrosky, Paul and Georges Rey. 1995. "When Other Things Aren't Equal: Saving *Ceteris Paribus* Laws from Vacuity". *British Journal for the Philosophy of Science* 46: 81–110.
Quine, Willard Van Orman. 1956. "Quantifiers and Propositional Attitudes". *Journal of Philosophy* 53 (5): 177–187.
Schiffer, Stephen. 1978. "The Basis of Reference". *Erkenntnis* 13: 171–206.
Segal, Gabriel. 1989. "The Return of the Individual". *Mind* 98: 39–57.
Sosa, Ernest and Pastin Mark. 1981. "A Rejoinder on Actions and *De Re* Belief". *Canadian Journal of Philosophy* 11: 735–9.
Stalnaker, Robert. 1984. *Inquiry*. Cambridge, (MA): MIT Press.
Tomkow, Terrance. 1992. *Against Representation*. Unpublished.

María Caamaño-Alegre

The Role of Presuppositions and Default Implicatures in Framing Effects

Abstract. Framing effects have hardly been studied from the philosophy of language. The variations in how subjects respond to positively or negatively framed descriptions of the same issue have received attention from social science research, where, nevertheless, a naïve understanding of speech interpretation has undermined the different explanations offered. The present paper explores the semantic-pragmatic side of framing effects and provides an explanation of this phenomenon in terms of pragmatic presuppositions and default implicatures. It is argued that the problem of valence framing includes two overlapping phenomena; on the pollster's side, there are wrong pragmatic presuppositions as to the kind of context that is relevant for survey interpretation, whereas the addressee proceeds by automatically connecting a certain kind of frame to a certain kind of implicit information related to the most common context of use.

1 Introduction

Framing effects are a widely studied phenomenon in social sciences, commonly understood as variations in how subjects respond to different but objectively equivalent descriptions of the same issue. As empirical phenomena, framing effects have been established to a very high degree of reliability and robustness (Kuhberger 1998). On the theoretical side, however, they are highly controversial since they challenge a common assumption in economic methodology, the one known as the 'principle of extensionality or invariance principle'. This principle says that individuals' preferences should not be affected by variations in the description of a problem.

Now, within the field of economic methodology, there are two conflicting ways of understanding framing effects; they can be regarded either as manifestations of our cognitive limitations, or as manifestations of our cognitive strategies. Only very recently a few attempts have been made at exploring the semantic-pragmatic side of this phenomenon. The present paper is intended to provide a contribution in this direction. I will restrict my analysis to the so called 'va-

María Caamaño-Alegre, Universidad de Valladolid, Contact: mcaamano2@gmail.com

https://doi.org/10.1515/9783110702286-011

lence framing effects',[1] i.e. effects caused by frames where the same issue is described either in positive or negative terms.

In the field of social sciences, the use of a wide variety of linguistic means to gather information about the subjects' beliefs, expectations, assessments or plans of action has significantly increased. In this respect, the reliance on various kinds of surveys and interviews has extended substantially, and faces the difficulties concerning the so-called 'framing effects'. Broadly speaking, these effects are related to the influence that different ways of presenting the same issue may bear on the respondent's response. The aim of this paper is to elucidate the semantic-pragmatic side of this problem, and, in particular, to explain how framing effects may be related to presuppositions and implicatures—this being a question that has been hardly dealt with in the standard literature on the subject. It is argued that different frames generate different inferential contexts by means of/through well-established linguistic practices.

The present account also challenges a common assumption in the sphere of economic methodology, usually referred to as 'the principle of extensionality' or 'the invariance principle' (Bourgeois-Gironde & Giraud 2009: 385 – 87), which establishes that individuals' preferences should not be affected by variations in the description of a problem. It is thus assumed that different ways of presenting the same set of possible options should thus not change the subjects' choices with respect to those options. Although behavioral economists have indeed diverged from the prevailing view in economics—arguing that framing effects should be approached not as mere cognitive flaws in the recognition of identical options, but as signs of the subjects' attitudes toward different aspects involved in those options—the explanatory factors identified by them fail to capture the importance of some semantic-pragmatic elements involved in the interpretation of frames. The influential studies by A. Tversky and D. Kahneman (1981, 1991) certainly shed some light on the way individuals process information depending on how the latter is presented to them. Although they did that mainly by empirically ascertaining several psychological biases—like loss aversion and the endowment effect, which are activated according to the kind of frame being used—they also acknowledged that the reference point regarding the value of an outcome does not stay neutral but varies depending on what is induced by the frame itself. The underlying semantic and pragmatic nature of this variation, however, is not analyzed by these authors.

After briefly characterizing framing effects and commenting on their standard explanation—which heavily relies on psychological loss aversion or the en-

1 In what follows I will talk of 'framing effects' when referring to valence framing effects.

dowment effect (section 2)—I will discuss the few attempts (Bourgeois-Gironde & Giraud 2009: 385 – 87, Moscati 2012: 8) at providing a semantic-pragmatic explanation of them in terms of situated linguistic understanding and a revised notion of extensionality (section 3). The remaining sections are devoted to examine the nature of the implicit information conveyed by frames and explore its understanding in terms of presuppositions and implicatures (section 4). I will argue that alternative frames trigger different interpretations regarding the most likely context of use of the frame. Finally, I will conclude that the notion of default implicature best explains the way different information is conveyed by alternative frames.

2 Framing Effects

As soon as the late 1990s, Levin, Schneider and Gaeth (1998) urged researchers to refine the typology of framing effects so that it became possible to account for the apparently inconsistent results achieved when trying to detect such effects. The plurality of interventions, moreover, entails a corresponding plurality of framing effects whose treatment requires equally differentiated procedures. In the typology suggested by Levin *et. al.*, three main kinds of valence framing effects are distinguished: the extensively discussed risky choice framing effect, and two other effects often overseen or mistaken for the latter, namely, attribute framing and goal framing. As explained by the authors (1998: 151, 181), each frame differs from the others in what is framed, what the frame affects, and how the effect is measured.

In the risky choice framing, the complete set of outcomes from a potential choice involving options with different levels of risk is described either in a positive or in a negative way. The framing effect is here measured comparing the rate of choices for risky options in each frame condition. Risk aversion would explain the fact that, when presented in negative terms, the riskier option is chosen by respondents more often than the safer one. A wide variety of experiments on risky choice,[2] from bargain situations to medical treatments, shows that when the outcome is described in terms of gains (lives saved, earned income), subjects' tendency to take risks diminishes. By contrast, such tendency increases when outcomes are expressed in terms of losses (lost lives, incurred debts). The paradigmatic case of risky choice framing effect is illustrated by the so-called "Asian disease

2 See Levin *et. al.* (1998: 154 – 157) for a collection of experimental results obtained within the domain of risky choice framing effects.

problem" (Tversky & Kahneman 1981). In this task, the two equivalent pairs of independent options with different level of risk are the following: a) a sure saving of one-third of the lives versus a one-third chance of saving all the lives and a two-thirds chance of saving no lives; b) a sure loss of two-thirds of the lives versus a one-third chance of losing no lives and a two-thirds chance of losing all the lives. The majority of subjects select the first option in the positively framed version of the task, and the second option in the negatively framed version.

In the form of framing called 'attribute framing', the positive or negative description of some characteristic of an object or event affects item evaluation, which is estimated by comparing the attractiveness ratings for the single item in each frame condition. The associative processes based on valence explain that positively described objects or events are more positively valued. This result has been established with much higher reliability and robustness than the other two kinds of framing effects compared by Levin *et. al.* (1998: 160). The fact that evaluations vary as a result of positive or negative framing manipulation has been established for issues as diverse as consumer products, job placement programs, medical treatments, industry project teams, students' level of achievement or the performance of basketball players.[3] Ground beef, for example, was rated as better tasting and less greasy when it was described as 75% lean rather than as 25% fat. Similarly, students' performance was rated higher when their scores were expressed in terms of percentage correct or percentage incorrect. Analogous results were obtained in the rest of cases.

Finally, in the case of goal framing, the same consequences of a conduct are specified either in positive or negative terms. The positive frame focuses attention on the goal of obtaining the positive consequence (or gain) associated with a given behavior, whereas the negative frame focuses attention on avoiding the negative consequence (or loss) associated with not performing such behavior. The variation in how persuaded an agent is to make or not make the decision to perform a certain conduct is regarded as an effect of the variations in the frames applied. The effect itself is measured by comparing the rate of adoption of such conduct under each frame condition. Experimental evidence shows that the negatively framed message, that is, the one emphasizing avoidable losses, proves more persuasive than the same message framed positively, and therefore stressing the potential gains. Real examples where goal frames are at use can be found in studies on the promotion of health, on endowment or on social dilemmas. Most subjects appear more inclined to adopt a certain conduct—like for ex-

3 See also Levin *et. al.* (1998: 161–163) for a lengthy compilation of experimental results related to attribute framing effects.

ample breast self-examination, use of public resources or of a credit card—when they receive information stressing the potential losses derived from not engaging in such conduct than when presented with information highlighting the potential profits resulting from engaging in it.

In the above-mentioned examples, individuals show to be more persuaded to adopt a given behavior when descriptions emphasize, respectively, the decrease in the probability of detecting cancer if no self-examination is carried out versus the increase of such probability in case a self-examination is performed, the losses suffered by the individual who contributes to the public goods versus the foreseen gains if the individual contributed to them, and the losses due to not using the credit card versus the benefits derived from its use.[4]

Despite the growing interest raised by the problem of framing effects, the majority of studies on these effects are focused on their diagnosis, while the attempts at explaining and controlling them are still extremely tentative and fragmentary. As already pointed out by Tversky and Kahneman (1981, 1991) in several of their influential studies on framing effects, the task of devising frames must be done by taking into account individuals' susceptibility to changes in reference points or in what is perceived as the *status quo* regarding some issue. Different frames would lead to different choices of reference points and, consequently, to a different way to encode the outcomes as gains or losses, which accordingly would bring about a different selection of options.

In discussing some of the anomalies affecting their prospect theory (Kahneman & Tversky 1979), both authors appeal to the possible occurrence of most likely intertwined phenomena like loss aversion and the endowment effect. These phenomena would emerge in most cases due to some framing conditions in which the reference point regarding the value of an outcome does not stay neutral but varies depending on what is induced by the frame itself. Let us recall that prospect theory, as opposed to classical theory, is committed to the view that risk aversion is dependent on a reference point. Under that assumption, it is predicted that risk aversion is linked to the domain of gains, and risk seeking to the domain of losses. In their 1979 paper, Kahneman and Tversky established that the above tendency could be reversed depending on the framing employed for the same pair of options. An initial remark in that direction can be found in some of their comments on the isolation effect (1979: 271), that is, individuals' inclination to ignore those components shared by alternatives and to focus on those making them different. Since there is more than one way to decompose a pair of alternatives into shared and distinctive components, the different

4 The wide range of real cases collected by Levin *et. al.* (1998) can be found in 169 – 171.

ways of decomposition may also prompt different preferences. This point is made more explicit as both authors refer to the reference point assumed by individuals and identifiable with those individuals' *status quo* or current state:

> "However, the location of the reference point, and the consequent coding of outcomes as gains or losses, can be affected by the formulation of the offered prospects, and by the expectations of the decision maker" (1979: 274).

Kahneman and Tversky go into great detail as to how reference points may vary, emphasizing that those reference points fixed by the *status quo* may shift as a result of encoding losses and gains relative to expectations that differ from the ones determined by the *status quo*. They also mention more specific cases where different encodings of the same pair of options create discrepancies between the reference point and the actual situation. According to them, this is exactly what happens when the choice is encoded in terms of final outcomes, as suggested from decision theory, instead of in terms of losses and gains (1979: 286 – 87).

In addition to loss aversion, endowment, preservation of the *status quo* and the tendency to ignore similarities, Levin and his collaborators point to the activation of positive associations in memory as the main mechanism responsible for attribute framing effects (1998: 164 – 5). Positive stimuli generated by a frame would yield some associative responses that, in turn, would cause a clear increase in the level of approval that each individual assigns to the positively described option as opposed to that assigned to the negatively described one. It has even been demonstrated that the mere activation of positive associations with respect to one of the options presented for a given choice brings about substantial positive distortions of that option against the other one (Russo, Medvec, & Meloy 1996: 103 – 107).

Turning now to the attempts at explaining goal framing effects, it is worth stressing the strong empirical support for the hypothesis of the negativity bias (Taylor 1991: 68 – 71). According to this hypothesis, individuals pay more attention to negative information than to equivalent positive information, showing themselves more influenced by the former than the latter. Since the 1990s, some of the explanations for the different framing effects have been partially unified; more specifically, loss aversion is now understood as a subclass of the negativity bias, and the *status quo* bias is in turn regarded as a subclass of the loss aversion bias. In all these cases, the rejection caused by a loss is higher than the desire to obtain a gain of the same magnitude (Levin *et. al*. 1998: 177).

To summarize, we have seen so far how risky choice framing effects have been explained on the basis of loss aversion, endowment, preservation of the *status quo* and the tendency to ignore similarities (Kahneman & Tversky 1979),

how attribute framing effects have been accounted for in terms of associative responses and selective attention mechanisms (Russo, Medvec, & Meloy 1996: 103 – 107), and how negativity bias has been emphasized as the main factor behind goal framing effects (Taylor 1991: 68 – 71). It is also worth examining those empirical findings pointing to variables that diminish or prevent such effects. In the case of risky choice framing, for example, it was demonstrated that when some question about the subject's reasons for a certain choice was added to the survey, then the framing effect was diminished or even eliminated. It is what Larrick, Smith and Yates call "the reflection effect" (1992: 199), which, according to their results, would make it possible to reverse framing effects by means of reflection on the issue presented within the frame. In a similar vein, Smith and Levin experimentally showed that individuals with a lower need for cognition were more affected by framing effects than those with a higher need for cognition: framing effects were hardly noticeable among the latter (Smith & Levin 1996: 283). As is well known in the field of psychology, the need for cognition constitutes a personality variable reflecting the individuals' disposition to perform cognitive tasks that require effort.

Experimental results suggest that factors other than the above also have a bearing on the scope of framing effects. These factors include the domain of problems presented, the traits of the experimental subjects, the magnitude or probability of potential outcomes, and the categories applied in verbalizing such outcomes (Levin *et. al.* 1998: 153). For instance, subjects are more inclined to take risks related to health issues than related to finances. The other two cases mentioned above, however, could be covered by the general case where the amount of information handled by the subject is inversely proportional to the scope of the framing effects (Schoorman *et. al.* 1994: 520). Notice that, as already commented, the variations in such amount may be due to variations intrinsic to the frame, and basically dependent on how detailed the frame is, or to variations in the subjects, mainly related to their need for cognition or degree of competence in the kind of subject presented. With respect to the traits of the experimental subjects, it has been found that experts or students in a certain field tend to be less affected by framing effects when confronted with options evaluable from such field. Similarly, it has been verified that replacing expressions like 'many' or 'few' with numerical values lowers the intensity of framing effects. In the study by Schoorman *et. al.* mentioned earlier, it has been experimentally established that the subject's degree of involvement or responsibility concerning a given issue can also eliminate the bias produced by the framing of the issue.

The situation is somehow different in the case of the bias caused by the attribute frame, for, as noted earlier, the sort of effect produced by this frame is the most homogeneous and clearly verified among ones caused by the valence frames.

Thus, despite the different domains of problems or the differences between subjects, the positive description of an item attribute, as opposed to its negative description, will almost always favor the more positive evaluation of both the attribute and the corresponding item. However, also in the case of attribute framing, a lower intensity of the bias has been experimentally determined when there is, on the subjects' side, a low degree of involvement as to the issue described (Marteau 1989: 90 – 93, Millar & Millar 2000: 860 – 63). We find here again a phenomenon that suggests an inverse relationship between the intensity of the framing bias and the level of processing of information provided to the subject. Therefore, this phenomenon might support the hypothesis, backed up by the experimental work of Maheswaran and Meyers-Levy (1990: 365), according to which the more involved an experimental subject is in the issue described, the more detailed their processing of the information related to the issue. Moreover, several experimental studies have shown the occurrence of a closely related phenomenon, namely, that the evaluation of real items is less affected by framing bias than the evaluation of hypothetical items. Attribute framing effects are also diminished when subjects are asked to explain their answers or give reasons for them.

As seen in the former cases, the degree of involvement in the topic presented, together with the tendency of the subjects to make a cognitive effort, are inversely related to the intensity of goal framing effects.[5] Perhaps because of the greater structural complexity of goal framing, there are more variations in operationalizing this framing, which ultimately entails a less homogeneous evidence for goal framing than for attribute framing (Levin *et. al.* 1998: 176). More specifically, such operationalization can be done either through simple negation (not obtaining profits) or through alternative terminology (losing the possibility of obtaining profits). Even if it seems obvious that linguistic variation may influence the strength of all sorts of valence framing effects, there are more potential linguistic variations in the case of goal framing, since the latter involves describing the consequences ascribed to some behavior as opposed to those ascribed to not performing such behavior. As Levin and his co-workers emphasize, in order to clarify when the responses of the subjects are dependent on semantic variations, it is necessary to develop an empirical study on language itself (1998: 174).[6]

5 Numerous references to empirical studies that point to this issue can be found in Levin *et. al.* (1998: 174).

6 In his 1992 paper, Rolf Mayer provides some clues to develop the kind of study suggested above. There he refers to some semantic aspects relevant in framing effects, such as the clustered nature of meaning, the impact of thematic roles or the distinction between discursive background and discursive front.

Despite the seeming diversity of explanations for framing effects, there are some explanatory variables shared by all of them. These explanations, in fact, all point to a common basic phenomenon, i.e. the fact that the rejection caused by a loss is higher than the desire to obtain a gain of the same magnitude. Moreover, in all cases the main explanatory variables associated to this phenomenon are:

a) the negativity bias, that is, the tendency to pay more attention to negative information than positive information (Taylor 1991: 68–71), which includes loss aversion and preservation of the *status quo* (Kahneman & Tversky 1979); and
b) the grasp or inference of implicit information about reference points (Kahneman & Tversky 1979), which concerns the implicit standard that is used in assessing the value of a potential gain or loss.[7]

Standard explanations of framing effects, however, face some serious limitations. There are two main problems undermining these explanations. On the one hand, psychological variables like loss aversion seem unsuited to explain response shifts in cases where the framed options are related to neither risk nor possible gains or losses, e.g. in cases of attribute framing of issues like the performance of basketball players or students' level of achievement. On the other hand, a naive understanding of speech interpretation seems to be underlying the assumption of the principle of invariance. To overcome these shortcomings, I am going to focus on what I regard as the most promising approach to the problem of framing effects, which entails exploring the connection between information and framing. Surprisingly this side of the problem has not been as carefully examined as its psychological side. Yet, as already noted, if we consider the relation between information and framing we find empirical evidence of an inverse relationship between the intensity of the framing bias and the amount of information provided to the subject, or the level of processing of such information (Schoorman *et. al.* 1994). For instance, using numerical values instead of natural language quantification or adding to the survey some questions about the subject's reasons for a certain choice have been proven to diminish the corresponding framing effects. These phenomena suggest that when information is not provided by the frame, addressees 'complete' such information—and they do that in a way unexpected by the pollsters. In parallel to this empirical evidence, there is an increasing acknowledgment of differences in the implicit choice-relevant information conveyed by frames. A new emphasis on choice-relevant information-

7 Reference points, in this sense, can be also understood as implicit assumptions about the *status quo*.

al equivalence as opposed to mere extensional equivalence between frames has been made accordingly. Against this background, the need to examine language and linguistic practices involved in frames becomes more and more clear.

3 Earlier Attempts at Explaining Framing Effects In Semantic-pragmatic Terms

There have been a few attempts at explaining framing effects in general on the basis of the traditional semantic distinction between extension (what is designated by an expression) and intension (the way of determining extension), which all questioned the way the principle of extensionality is usually understood or applied.

In the field of philosophy of economics, for example, Moscati has recently argued for understanding framing effects as doxastic effects caused by the intensional discrepancy between extensionally identical descriptions. Surveys employed by Tversky and Kahneman in their experiments included extensionally equivalent descriptions of outcomes and probabilities which, nevertheless, intensionally differed by virtue of the way uncertainty was presented, either in one-stage games or in two-stage games (Moscati 2012: 7). Moscati points to the problem of referential opacity in intensional contexts as what would explain the apparent irrationality of subjects' choice reversals (2012: 8). According to this author, surveys constitute intensional contexts where descriptions are interpreted as tied to beliefs. The apparent manifestations of irrationality would then be the consequence of an apparent co-extensionality, mistakenly taken as real by those researchers who overlook the opaque nature of intensional contexts such as that of subjects' beliefs. He conceives framing effects more generally, "as the effects, on beliefs, preferences or decisions, of intensionally different descriptions of an extensionally single object" (Moscati 2012: 7–8). This framework highlights the economic relevance of interactive beliefs and interactive knowledge—that is, respectively, beliefs or knowledge that an individual has about what other individuals believe or know about the world—i since in many cases individuals take action on their basis (Moscati 2012: 14).

Other authors like McKenzie and Nelson (2003) or Bourgeois-Gironde and Giraud (2009) develop a similar approach in claiming that frames bring with them a leakage of choice-relevant information about the speaker's reference point. Therefore, rather than being objectively equivalent, alternative frames would leak information allowing one to infer the existence of certain background condition from the speaker's choice of frame. Bourgeois-Gironde and Giraud

(2009: 385–87) make use of the distinction between intension and extension with the purpose of explaining the mechanism by which framing effects come to happen. This distinction would make it possible to account for those cases where different descriptions of the same problem prompt different choices. Both authors draw attention to the fact that, in economic methodology, the principle of invariance or extensionality goes beyond the logical principle, establishing the co-extensionality between expressions whenever the latter are interchangeable *salva veritate* (i.e., whenever the truth value is preserved). In the context survey research, what needs to be guaranteed by means of co-extensional descriptions is not only the truth value preservation but also the preservation of whatever information is relevant for making decisions. What needs to be specified, therefore, is the kind of information regarded as relevant for purposes of deciding among the options presented. Only after such information had been specified could framing effects be ascertained as violations of the extensionality principle in the contexts of decision under study. Violating extensionality would then imply that choice-irrelevant information determines the choices or judgments made by the subjects.

'Intension', however, is used by these authors in a sense that may include explicit contents (conventional meaning, truth conditions) as well as implicit contents (speaker's meaning, contextual information). Unlike the standard notion of intension, usually restricted to explicit contents, this broad notion is tightly associated with implicit contents, whose nature, however, remains highly underdetermined. In fact, the emphasis should be placed only on implicit contents, since the intensional aspect has traditionally been equated with explicit contents. Furthermore, if we grant that alternative valence frames are usually designed not only to guarantee interchangeability *salva veritate*, but also so that they share the same explicit contents, then response shifts induced by alternative frames must be due to differences in their implicit contents. To be clear, when the above authors talk about intension, most likely they are pointing to implicit contents related to doxastic elements attributable to the speaker by the addressee.

Once implicit contents are brought to the foreground, two related questions arise: a) what is the nature of the implicit information conveyed by extensionally equivalent frames sharing the same explicit contents? And b) how is the implicit information conveyed by the frames? Drawing on some empirical data presented below, my suggestion to answer the first question is that the (choice-relevant) implicit information conveyed by the frame is about the most likely context of use of a frame, or, to be more precise, about some typical background conditions corresponding to such context. This information is not part of what is asserted in the frame, but rather part of what is assumed about the context whenever a certain

frame is employed. The resulting assumption concerns neither the intentions of any particular speaker, nor any other particular contextual aspect surrounding the framed utterance, since surveys are usually non-conversational contexts where both the 'speaker' (pollster) and the framed issue are absent. In implicitly conveying contextual information typically associated with a frame, valence framing induces an addition of a proposition to the proposition expressed by an utterance, which brings us to the second question. The addition induced by framing seems to occur through a process of standardization, i. e. by way of a regular pattern of use or choice of a frame whenever certain contextual conditions have arisen. This is also suggested by recent empirical data on frame choice.

Going back to the views put forward by Moscati as well as Bourgeois-Gironde and Giraud, it should be noted that they all indirectly point to an essential feature of framing effects, namely, a doxastic (or knowledge) condition concerning familiarity with the usual background conditions of frame choice. Nevertheless, by placing the focus on the speaker's beliefs about reference points rather than on the typical contextual conditions determining frame choice, they fail to capture the ultimate nature of the phenomenon under study. Consequently, they wrongly assume that addressees make inferences from the use of a frame to the speakers' beliefs and from them to certain background conditions obtaining. The process of standardization governing framing effects, however, suggests that the addressees' inferences are made from the use of a frame directly to certain background conditions obtaining.

So far, I have only mentioned some basic features of framing effects that have not been properly discussed in the literature. However, I have not yet provided a framework within which those features could be explained and made more precise. In particular, I have not explored the different notions—like presupposition, implicature, default meaning, etc.—that could help to better understand the role that implicit content plays in framing effects. The main candidates to play this role will be examined in the next section.

4 Implicit Content and Framing Effects

Before examining the different available notions that could be applied to characterize the implicit content conveyed by frames, let us take a look at the empirical evidence supporting the view so far adopted here.

Some empirical data collected over the last decade show that listeners (or readers) are able to make inferences about current or presupposed states from the speaker's (pollster's) choice of frame (Sher & McKenzie 2006). Note that the inference is from a choice of frame to a presupposed state, not from an utterance to a

belief/intention and from the latter to a presupposed state. In some of the cases studied, depending on whether the glass was described as half empty or half full, readers were able to infer its previous volume of liquid (the inference being that the glass was previously completely full or completely empty, respectively).

A hypothesis that has been developed for natural language quantification—but which, as argued by Moxey, could be generalized to valence framing—is that focus and polarity together are the main kind of presupposition trigger (Ingram 2010, Moxey 2011). The intuitive idea is that negation is most informative if interpreted as a denial of a positive alternative (i.e. a complement set) and *vice versa*. Focus can thus be originated by a choice between alternative frames, thereby yielding a soft presupposition (or assumption) trigger regarding the existence of a complement set—a full glass as opposed to a half empty glass or the reverse. Valence framing leaks information about a complement set that is usually part of the objective context when a reference set is mentioned in a description (Ingram 2010, Moxey 2011). Empirical research on natural language quantification supports the claim that negative quantifiers (like 'not many' as opposed to 'a few') lead interpreters to infer that the small amount denoted is in contrast to a larger supposed amount. Conversely, terms like 'few' (Sher & McKenzie 2006) 'leak' information about a higher reference point. This shows that, as interpreters, we seek out information not only about what is in fact the case, but also about what is assumed about the context, especially if deemed choice-relevant. Consequently, choice of expression implicitly conveys information on the facts while conveying the facts themselves. This information is tightly connected to usual opinions or expectations on the facts in question and rooted in a standard choice of frame alternatives in certain contexts. As interpreters we seek out such information as part of what we understand given the writer's message.

Interestingly, Moxey has extended this 'presupposition denial account' of focusing properties of natural language quantification also to valence frames in general (2011: 122–3). The label for the account refers to the basic assumed fact that, in interpreting negation, we presuppose that it involves a denial of a positive alternative (i.e. a complement set) since this maximizes the information we can get from the utterance—by the same token, we presuppose that a positive frame involves a denial of a negative alternative. The polarity of natural language quantification serves to frame quantity information in either a positive or a negative way. Each quantifier activates a normal pattern of focus on a complement set relative to the reference set (the overall set would include both sets). Now, as Moxey illustrates with several examples, this very account can be generalized beyond the domain of natural language quantification. An utterance like 'John didn't eat a cheese sandwich' places in focus a positive alternative, i.e. that John might have eaten a cheese sandwich, but this state of affairs did not hap-

pen (Moxey 2011: 119 – 121). Focus can thus be originated by a choice between alternative frames, thereby yielding a soft presupposition trigger regarding the existence of a complement set—a full glass as opposed to a half empty glass or the reverse. Valence frames, therefore, leak information about a complement set that is usually part of the objective context when a reference set is mentioned in a description. Going back to one of our former examples, it becomes clear that, depending on how the reference set is described in a sentence (for instance, a piece of beef as being 75% lean), there is a focus on a complement set, i.e., on the assumed average qualities ascribed to the sort of thing included in the reference set (pieces of beef being usually less than 75% lean).

After summarizing the main empirical insights into addressees' assumptions based on the speakers' choice of frame, the next step is to identify the notion that best captures all the above features. The possible candidates should account for the following features of the phenomenon under study:

a) it involves an addition to the proposition explicitly expressed by the utterance (the proposition that the glass was empty before being half full is added to the proposition that the glass is half full);
b) the addition is part of the addressee's interpretation of the utterance, not part of what the pollster's meant by the latter (the addressee, not the pollster, assumes the glass was empty before being half full);
c) what is added concerns not current but typical contextual conditions associated with the use of a frame (there is no glass in the present context, but the utterance is interpreted by considering how the typical situation is at the times when that kind of utterance is framed in a certain way);
d) the addition is about a complement set relative to a reference set explicitly mentioned in the utterance (an empty glass relative to the half full glass);
e) the addition is triggered by a focus on a complement set, resulting in turn from a choice of a frame over the other alternative (focus on a glass being empty before being half full as a result of choosing the positive frame 'half full' over the alternative negative frame 'half empty');
f) the addition is automatic (as soon as the frame is identified the assumption about the previous state of the glass as being half full or half empty is made);
g) the addition is easily cancellable (if a description of the previous state of the glass as being full is explicitly added to the positively framed utterance about the half full glass, then the usual assumption that the glass was previously empty is cancelled).

Since the authors dealing with the empirical data talk about presupposition and lexical alternatives, first I am going to explore the possible application of these

two concepts and then I will consider implicature in general as well as default implicature in particular, all widely discussed notions that have proven relevant in understanding the nature of implicit contents. On the other hand, given our present purpose, the details and debatable points in the analysis of these notions are not going to be tackled here. Instead, I will only consider a schematic and rather uncontroversial version of them in order to see whether they can accommodate and shed some light on the above phenomenon.

4.1 Exploring the role of presuppositions in framing effects

Let us start by considering presupposition. It is commonly understood that one sentence presupposes another whenever the second is true regardless of the truth or falsity of the first. That is to say, a presupposition projected from sentence *s* is also projected under negation ¬*s*. 'The present king of France is happy' presupposes the proposition that there is a king of France, which is triggered by the definite description included in the sentence. Since assumptions triggered by frames are not projected under negation, they do not fit this notion of presupposition. From the sentence 'the glass is not half full', we would not assume that it was previously empty; in all likelihood we would not know what to think about the state of the glass prior to not being half full. However, it is customary to distinguish between a semantic conception of presupposition and a pragmatic one (Simons 2013, Potts 2015). Semantic presuppositions would be linguistically triggered by some lexical item—like the definite description construction '*the-noun-phrase/singular common noun*' in the example just mentioned. They are necessary to determine the truth conditions of the sentence projecting them, which entails that, whenever a presupposition projected from a sentence is false, the sentence is not truth-evaluable—again like in the example. Assumptions triggered by frames are clearly not necessarily involved in determining their truth value of framed sentences, or, to put it differently, the truth of the latter can be evaluated without taking into account the assumptions that they convey.

It is possible to find a more promising approach to the problem, then, if we move on to pragmatic presuppositions. After all, according to the Stalnakerian picture, they are not primarily projected from sentences (not even from generic uses of sentences) but from the agent's conversational dynamics (Simons 2013: 7). He characterizes pragmatic presuppositions as the agent's beliefs about common ground (Stalnaker 1974, 2002: 716), i. e. about common beliefs regarding what propositions are accepted by all parties in a conversation. To put it more intuitively, presuppositions could be equated with beliefs about what is

taken for granted in a conversation, and therefore about the background of beliefs shared by the interlocutors or the background of propositions treated by them as true for some reason. The hearer's identification of a speaker's presuppositions would thus require the identification of the latter's intentions and beliefs in a conversational context. Simons' example of a contextual presupposition would be a case in point; if the chair of a meeting, which is supposed to start at 3:00, says to the audience 'OK, it's 3 o'clock', hearers would assume that it is time to start. In this case there is a complete proposition that is literally expressed and has nothing to do with starting the meeting—the fact that it is 3 o'clock—and something is added to this, namely, the proposition that it is time for the meeting to start, which constitutes a presupposition projected from the speaker's conversational dynamics. This addition is indeed not required to determine the truth value of the explicitly expressed proposition. As a consequence, presupposition failure would not result in the truth-non-evaluability of such proposition. Note, moreover, that these pragmatic contextual presuppositions are not required to pass the negation test, that is, if the chair said 'OK, it isn't 3 o'clock', the assumption would no longer be that it is time for the meeting to start. So, it seems that the main difference between presupposition and conversational implicature vanishes. In so far as the notions of pragmatic presupposition and conversational implicature merge, all the considerations below are applicable to both.

The question is whether the focus on a complement set originated by a choice between alternative frames is such as to trigger a wrong pragmatic presupposition on the addressee's side regarding the pollster's beliefs about common ground.[8] One essential aspect of Stalnaker's notion of pragmatic presuppo-

8 In discussing lexical alternatives as a source of pragmatic presuppositions, Dorit Abusch (2002: 8–11) argues that focus introduces an alternative set, which is turned into a pragmatic existential presupposition if the speaker pragmatically presupposes that some alternative is true. Although this view may seem at first promisingly close to the explanatory scheme I am looking for, there are important differences between the approach endorsed by Abusch and the one that Ingram, Moxey, Sher and McKenzie suggest. First, in alternative semantics, alternatives are considered part of the linguistic meaning (Fălăuş 2013); more precisely, they are seen as lexically encoded opposites ('stop' would encode 'continue' as its alternative). By contrast, in the presupposition denial account, the complement set is neither part of the linguistic meaning of the utterance nor understood as an opposite to the reference set. Second, focus is understood only as an intonationally prompted phenomenon according to alternative semantics, while in the presupposition denial account it is regarded as dependent on the choice of frame. Third, in frames, the kind of pragmatic presupposition triggered by focus is not about the existence of an alternative set (i.e. an opposite) but about the existence of a complement set (i.e. not an opposite but a standard contrast class). There is no single obvious opposite for a half full

sition is its emphasis on the importance of identifying the speaker's intentions and beliefs (1974, 2002), and it is this very aspect that makes it difficult to accommodate the kind of presupposition triggered by valence frames. The sort of framing used in surveys operates in non-conversational textual contexts where there is no speaker. In order to overcome this difficulty, the modified notion of pragmatic presupposition introduced by Marina Sbisà (1999: 330), explicitly developed to be applicable to text understanding, may prove useful. She argues that pragmatic presuppositions are shared beliefs about the objective context rather than about others' representations of objective context. Shared beliefs would be the result rather than the essence of common ground. One of the main ideas behind her view is that not only speakers carry pragmatic presuppositions, but also sentences do. Her account of pragmatic presupposition is developed to be applicable to text understanding, and so it proves relevant for generic non-conversational written contexts where the speaker is absent, like in the case of surveys. Beliefs about objective context could thus be understood as including beliefs about background conditions involved in framing effects.

We could try to reconcile Stalnaker's and Sbisà's views by arguing that the common ground involved in framing may be more complex than usually thought and include assumptions not only about the others' beliefs concerning some implicit information that is taken for granted (for instance, '25% fat' being equivalent to '75% lean') but also about what conditions of the objective context make it more appropriate to use one frame rather than the other (average level of fat being usually under 25% makes it more appropriate to use '25% fat' instead of '75% lean'). Now, if we decide to go down this road, we should be able to explain why common ground is not shared in survey contexts, that is, we would have to account for the fact that pollsters and addressees consistently hold different assumptions concerning the appropriate conditions for using one frame over another in survey contexts, even though they all take for granted the equivalence between '25% fat' and '75 lean'. Notice that all that is required for those assumptions is a competence as users of frames (knowing how), not a propositional knowledge of the competence (knowing that).

So we can arrive at the following explanation of framing effects in terms of pragmatic presuppositions: where pollsters presuppose that, in a survey context, describing a piece of beef as being '75% lean' is equivalent to describing it as being '25% fat', respondents take it as stressing that percentage over the

glass, since 'a not half full glass' ambiguously suggests many different alternatives to "a half full glass", 'an empty glass' being only one of them. Thus, despite employing similar terminology, alternative semantics and the presupposition denial account of framing effects diverge in important ways.

usual percentage, which would be presupposed to be lower than 75%. The disagreement arises, then, because (within survey contexts) pollsters do not endorse the respondents' assumptions regarding the relevance of both the usual percentage (below 75%) and the typical linguistic practice consisting in choosing a positive frame to stress a gain with respect to the average context (or a negative one to emphasize a loss with respect to the average context).

Two comments are in order here. First, none of the above assumptions are related to the speakers' intentions or beliefs, but rather to certain objective conditions concerning the *status quo* (the prevalent states of affairs), and to certain objective facts about linguistic practices typically related to those conditions. Second, pollsters do endorse such assumptions outside survey contexts, otherwise there would be no typical pattern of frame choice, and empirical evidence does confirm such pattern. Thus an interesting point is that the key explanatory variables are to be found in the objective contexts respectively connected to well-entrenched linguistic practices. There are facts about (frame) use that have a life of their own regardless of intentions on the part of the participants in a linguistic exchange. This is precisely why it is also difficult to explain the way that frame effects relate to defective contexts, i.e. contexts where some of the participants have false beliefs about common ground. When frame effects happen, there are false beliefs, neither about the objective context nor about the other participants' beliefs regarding such context, but only about whether the objective context, instead of an idealized context, is the one to be taken into account to interpret the sentence.[9]

After all, the main disagreement between pollsters and addressees lies in how they represent the context by default in surveys: pollsters assume an idealized context whereas addressees assume the most likely context of use.[10] Frames

9 This problem has also being pointed out by Jones in his Gricean analysis of economic experiments: "This means that it may be that some aspect of the experiment reminds subjects of a norm existing outside the experimental laboratory. The experimenter did not intend for this norm to be followed and so a systematic misunderstanding by the subject may occur as a result" (Jones 2007: 171).

10 This conclusion is in tune with Norbert Schwarz's view on 'the logic of conversation' in questionnaire research: "(...) research participants go beyond the literal meaning of the information provided by the researcher and draw on the pragmatic rather than the semantic meaning of the researcher's contributions. The researcher, however, evaluates participants' judgments against a normative model that draws only on the logical implications of semantic meaning at the expense of the pragmatic implications of the researcher's utterances" (Schwarz 1996: 7). His approach, however, mainly provides a Gricean analysis of research communication, showing that conversational norms (and, thus, implicatures) influence question interpretation (Schwarz 1996: 15–16). As argued below, although the Gricean framework explains some general aspects

operate in a peculiar, more systematic, linguistically marked fashion than suggested by the traditional notion of pragmatic presupposition. *Pace* Stalnaker, the key point here does not lie in what beliefs or intentions we attribute to the actual speaker/pollster, but rather in what typical contextual information emerges when a certain frame is chosen over other options. The speaker's actual beliefs lose importance in comparison to standard or well-entrenched uses of different linguistic frames.

The common ground involved in framing effects includes at least two assumptions, on the side of the respondents, regarding the objective context. One is about reverse properties (25% fat being the reverse property of 75% lean), whereas the other assumption concerns what prevalent conditions (or *status quo*) of the objective context make it more appropriate to use one frame rather than the other (average level of fat being usually under 25% makes it more appropriate to use '25% fat' instead of '75% lean'). It is important to stress that pollsters do endorse these assumptions outside survey contexts, for empirical evidence confirms the typical patterns of frame choice.

4.2 Implicatures and framing effects

It could be argued that the intuition behind this account of presuppositions involving well-entrenched or crystallized uses—i.e. the fact that negation is most informative if interpreted as a denial of a positive alternative and *vice versa*—nicely fits Gricean maxims of quantity and relation,[11] and that the inference prompted by frames can be better accommodated by applying the notion of *generalized conversational implicature* (Grice 1975) rather than the idea of pragmatic presupposition. Implicatures are inferences in which the inferred proposition bears no truth functional relation to the utterance contained in the text. They are taken to arise from the interaction of the proposition actually expressed in the utterance, certain features of the context, and the assumption that the speak-

involved in framing effects, some essential features of the latter are better captured by the notion of default implicature, which in turn is closely connected to that of default meaning. Note, moreover, that valence framing is not included in Schwarz's discussion of the formal features of questionnaires (1996: chapter 5). The same goes for Jones' proposal to apply the Gricean framework in order to clarify the notion of understanding involved in economic survey research and experiments (Jones 2007).

11 The first states that one should try to be as informative as one possibly can, and give as much information as is needed, and no more, the second, that one should try to be relevant, and say things that are pertinent to the discussion.

er is obeying the Cooperative Principle. In the case of generalized conversational implicatures, the inferences have 'crystallized' as a result of the standard use of a propositions expressed by the utterances, and so the context becomes irrelevant. An implicature of this kind is one which does not depend on particular features of the context, but is instead typically associated with the proposition expressed (in this case, with the frame chosen).

The notion of generalized conversational implicature may supplement the idea of pragmatic presupposition in a significant way, particularly in what concerns those additions connected to the typical use of a frame. Nevertheless, within the Gricean framework, the hearers' inferences are always primarily about the speakers' beliefs or intentions rather than about the objective context, and this would clearly be in conflict with my remarks made above. Furthermore, this is the same problem as the one already pointed out with regard to Stalnaker's notion of pragmatic presuppositions.

If, at least for the case of valence framing and text understanding, we relaxed both Grice's notion of conventional implicature and Stalnaker's notion of pragmatic presupposition so as to leave out the requirement concerning the content of the inferences, then framing effects could be explained on the basis of these notions as prompted by implicit information supplementing the information conveyed by the assertion. Such supplementary information would make it possible for the addressees to update their representation of the objective context. On the other hand, if we took a stricter view on presuppositions like Sbisà does (1999: 332–335), and were to understand presuppositions as assumptions (related to objective context) that ought to be shared and that should be taken for granted for an utterance to be acceptable, then it is not clear that framing effects could be accounted for in terms of pragmatic presuppositions. In this case we might rather have to rely instead on conversational implicatures, since, even when they are about the objective context and happen to be shared knowledge, they are not necessarily subject to any normative requirement concerning the representation of the objective context. To say that a piece of beef is '75% lean' when that percentage is below the average certainly does not make the utterance inacceptable, while—to use Sbisà's example—to say that 'John realized that he was in debt' when we know that John is not in debt does make the utterance unacceptable (1999: 334). It seems, however, that the very notion of acceptability involved here may allow for different, more or less strict characterizations. According to Stalnaker, for example: "a speech act is conversationally acceptable in the relevant sense just in case it can reasonably be expected to accomplish its purpose in the normal way in which the normal purposes of such speech acts are accomplished" (1974/1999: 51). Thus, it could be argued that describing a piece of beef as being '75% lean' when that percent-

age is below the average does after all violate 'the normal way' in which that frame is used for conversational purposes. Since the possible resolution of this dispute is clearly beyond the purpose of this paper, I will limit myself here to showing that, in cases where conventional implicatures are shared knowledge about the objective context, they retain the main intuitive features associated with pragmatic presuppositions.

All in all, the problem of valence framing is twofold, including two overlapping phenomena that create the 'perfect storm' conditions for survey interpretation to go astray. On the pollster's side, there are wrong presuppositions concerning the kind of context that the respondent will take into account in interpreting an utterance. Within survey contexts, pollsters operate with the idealized assumption that describing a piece of beef as being '75% lean' is equivalent to describing it as being '25% fat', and do not endorse the respondents' assumptions regarding the relevance of the typical linguistic practice consisting in choosing a positive frame to stress a gain with respect to the average context (or a negative one to emphasize the converse).[12]

The pollsters' mistake can be due to two different situations: a) they know the kind of default reasoning usually involved (when a certain frame is employed) but wrongly believe that the addressees will be able to identify the ideal nature of survey contexts and suspend such reasoning; b) they do not know what kind of default reasoning is usually involved (when a certain frame is employed) and wrongly believe that the ideal nature of survey contexts is common ground. Either way we have a defective context due to the pollster's wrong presupposition regarding (common ground on) the relevant context, although in a) that goes together with endorsing a wrong informative presupposition as to the possibility of changing the common ground in survey contexts so that respondents assume that the idealized context is the relevant one for interpreting the sentence. Informative presuppositions occur whenever a speaker utters a presupposing sentence perfectly knowing that the presuppositions of the sentence are not part of the common ground, but at the same time believing that they will be common ground following the utterance (Simons 2003: 16 – 20). So, a) describes the possible case where pollsters mistakenly presuppose that the idealized nature of the relevant context for interpreting an utterance will become common ground once the respondent receives the utterance in a survey context.

12 This kind of disagreement concerning presuppositions quite naturally fits the notion of soft trigger, i.e. an optional presupposition that can occur only when it fits into the context and can be easily suspendable (Abush 2002).

Addressees, by contrast, proceed in quite a different manner. We have explored the possibility that generalized conversational implicatures are essentially involved in their understanding of framed survey questions. However, there are some remaining problems that undermine the plausibility of this approach; in particular, the fact that the addition is automatic (as soon as the frame is identified, the corresponding assumption is made), and the fact that it arises locally (as soon as a construction reveals the kind of frame used, the addition is triggered).

As I will show in the next section, the notion that proves most useful in accounting for the above issues is default implicature. In connecting a certain kind of frame to a certain kind of implicit information, addressees operate by default interpretation, and such interpretation is not suspended in the survey context. To summarize, presupposition, whether informative or not, plays a role on the side of the pollster, and default interpretation does so on the side of the addressee.

4.3 Understanding framing effects in terms of default implicature

As emphasized by Katarzyna Jaszczolt (2014), despite the fact that there is no consensus as to how default interpretations should be understood, the notion of default meaning proves helpful in distinguishing between salient, automatic interpretations and costly pragmatic inferences. A generic characterization of default meaning is stated by her as follows: "(...) default interpretation of the speaker's utterance is normally understood to mean salient meaning intended by the speaker, or presumed by the addressee to have been intended, and recovered (a) without the help of inference from the speaker's intentions or (b) without conscious inferential process altogether" (Jaszczolt 2014).

Notwithstanding the reference to intended meaning in the first lines of the above definition, point a) immediately rules out inference from the speaker's intentions as an element playing any role in default interpretations. Both points a) and b) stress the context-independent nature of default interpretation, which becomes more obvious in cases where default meanings are clearly triggered by a construction—as happens in the case of framing effects. To use one of Jaszczolt's examples, we interpret the sentence 'Many people liked Peter Carey's new novel' as meaning, by default, 'Many, but not all people liked Peter Carey's new novel'. The same way that such interpretation is automatically and locally prompted by the construction 'many x', the sentence 'the glass is half empty' includes the neg-

ative frame construction 'half empty x', which locally triggers by default the interpretation 'the glass that was previously full is now half empty'.

As pointed out earlier, there is no agreement on whether context-independence, locality, cancellability and exclusion of conscious inference are essential properties of default interpretation. Consequently, there are also conflicting views on the semantic, pragmatic or even grammatical nature of default meanings. In order to avoid issues that would by far exceed the limits of the present discussion, I am going to draw attention solely to those aspects of default meanings that prove most relevant in understanding framing effects. What seems to be clear is that the very notion of default meaning calls for a recognition that utterance interpretation involves a variety of mechanisms, in the form of conventions and heuristics pertaining to rational communicative behavior. Addressees' shortcuts to meaning recovery constitute one of such mechanisms. These shortcuts, which are generated by a process of standardization in the use of a construction, are based on assumptions regarding both scenarios and human mental processes. In the cases where these shortcuts or defaults operate, context and inference may not play any significant role in the addressees' recovery of meaning (Jaszczolt 2010/2015: 744).

The characterization of default implicature that best accounts for the phenomenon of framing effects includes all the above-mentioned properties (context-independence, locality, cancellability), being rather close to Stephen Levinson's notion of presumptive meanings (Levinson 2000), which deviates slightly but significantly from Grice's notion of generalized conversational implicature, particularly as regards the features of locality and independence from the speaker's intended meaning. According to the Gricean picture, non-literal interpretations only occur after the addressee has grasped the literal meaning of the uttered sentence, i.e. they are a global phenomenon related to the overall explicit meaning of the sentence. Levinson argues, on the contrary, that some lexical constructions can locally and by themselves prompt non-literal interpretations by the addressees. The sentence 'Some boy came' is interpreted as 'Not all of the boys came' by virtue of it including the word 'some' that by itself leads to the interpretation 'not all' (Levinson 2000: 36–37). Analogously, the positive frame construction 'half full' by itself triggers the reading 'previously empty and now half full'. Also, negative frame constructions like '20% fat' or '20% errors' are understood, respectively, as expressing 'being 20% fat and above the average level of fat' and 'having 20% errors and being above the average levels of errors'. The same way that Levinson explains cases like 'some' by appealing to the Q-heuristic ('what isn't said, isn't'), we could appeal to the following heuristic for the case of frames: where a positive frame is chosen, it can be assumed that the positive property is above average—and the same goes for

negative frames. It is no coincidence that Levinson's reflections on the 'bottleneck of communication' closely resemble Kahneman's views on the practice of 'thinking fast' as opposed to 'thinking slow' (Kahneman 2011). Simultaneously minimizing usage of linguistic tools and maximizing meaning recovery leads to a poor linguistic articulation followed by a fast processing by the addressee. Hence, unsurprisingly, the easy and relatively frequent cancellability of local additions, a feature that has been noted by Levinson and that, as shown before, is also shared by local additions triggered by frames.

On the other hand, Jaszczolt's notion of defaults embraces two kinds of default meanings, i.e., cognitive defaults, triggered by the properties of human inferential system, and social, cultural and world-knowledge defaults, triggered by the shared background on social conventions and knowledge of both cultural and physical properties of the environment (Jaszczolt 2010/2015: 746 – 750).[13] These two sources of default meanings would automatically yield certain information whenever a certain construction is employed—or, if we endorsed Jaszczolt's wider account, whenever a certain typical situation occurs.[14] To use her own example, world-knowledge defaults would be responsible for interpreting 'and' as 'and as a result' in sentences like 'The temperature fell below -10 degrees Celsius and the lake froze'. As for inferential system defaults, they would explain the default referential as opposed to the attributive interpretation of definite de-

13 In Jaszczolt's view, these defaults could combine with other components of meaning such as knowledge of word meaning and sentence structure, knowledge of the situation of discourse and conscious pragmatic inferences (Jaszczolt 2010/2015: 750). Depending on whether the contribution of defaults and conscious pragmatic inferences is more or less salient, they could or could not be regarded as part of the explicit meaning. Since my goal here is not to develop a theory of meaning, but just to show the special usefulness of the notion of default meaning to explain framing effects, the merger representation of meaning is only mentioned for purposes of contextualization of her ideas.

14 A somehow striking consequence of applying default semantics in explaining framing effects is that added contents creating these effects might not be regarded as implicit contents but instead as explicit ones. From this approach, explicit contents are not primarily defined in terms of literal meaning, but in terms of salience (Jaszczolt 2010/2015: 743). To be clear, according to this view, given the salience of many instances of implicatures, they could be considered as providing the explicit meaning of the uttered sentence (Jaszczolt in 2010/2015: 745 – 6, 749 emphasizes the ample experimental evidence showing that the explicit meaning of a sentence often corresponds to implicatures). In this respect, default semantics diverges significantly from other Gricean approaches, which do keep the syntactic constraint on what is said. As a consequence, salient meanings are orthogonal to the explicit/implicit distinction. Since adherence to this feature of default semantics is not required to support the present analysis, a minimal notion of default meaning is here applied together with the traditional distinction between explicit and implicit contents.

scriptions like in 'The author of *Don Quixote* fought in the Battle of Lepanto' (interpreted as 'Cervantes fought in the Battle of Lepanto').

In the case of framing effects, the shared cultural background regarding standard uses of frames is definitely involved in triggering default meanings. Whether they are also cognitive defaults related, for instance, to the human tendency to operate with contrast classes and to project the past onto the present, is an interesting question that goes beyond the limits of the present paper. A straightforward relation between default interpretation and framing effects seems to emerge once we summarize the above points and retrieve the seven features of framing effects mentioned at the beginning of the section. The fact that framing effects involve an automatic, frame-triggered addition to the proposition expressed by the utterance (features a, d, f) clearly accords with the notion of default implicature. As soon as the frame is identified, without mediation of conscious inference or consideration of the context, an assumption about a complement set (relative to a reference set explicitly mentioned in the utterance) is triggered. The fact that, despite not being meant by the pollster, such implicature is made by the addressee on the basis of the standard conditions associated to the use of a frame (features b, c) further reinforces the presumptive, context-independent nature of frame interpretations. Moreover, both the source and the content of default interpretations involved in framing effects—that is, both the competence in frame choice and knowledge of usual background objective conditions concerning complement sets (features c, d, e)—suggest that at least some cultural and world-knowledge defaults play an essential role in such phenomenon. Finally, the easy cancelability of assumptions triggered by frames (feature g) clearly shows that, even if standardly connected to frames, the first should not be explained in terms of semantic presuppositions.

5 Conclusion

On the side of the addressee, framing effects result from default interpretations triggered by focus and polarity that in turn are generated by a choice of frame. This kind of interpretation, which concerns assumptions about objective background conditions for framing, is triggered by standardized, well-entrenched linguistic practices involving a certain choice of frame given some prevalent states of affairs. On the side of the pollster, the problem arises due to the pragmatic presuppositions assumed, within survey contexts, with regard to the relevant context for interpretation.

Acknowledgements

I am thankful to Agustín Vicente for helpful comments on earlier versions of this paper. Thanks also to the participants in the The first *Context, Cognition and Communication Conference Context dependence in language, action and cognition* (Warsaw 2016), where I had the chance to get a valuable feedback on different aspects of this work. Many thanks are also due to Emad Ali Khan for correcting my English.

This research was financially supported by the research projects "Laws and Models in Physical, Chemical, Biological, and Social Sciences" (PICT – 2018 – 03454, ANPCyT, Argentina), and "Stochastic Representations in the Natural Sciences: Conceptual Foundations and Applications (STOCREP)" (PGC2018 – 099423 – B – I00, Spanish Ministry of Science, Innovation and Universities).

References

Abusch, Dorit. 2002. "Lexical Alternatives as a Source of Pragmatic Presuppositions". In Brendan Jackson (ed.), *Semantics and Linguistic Theory* (*SALT) XII*. Ithaca NY: CLC Publications. http://conf.ling.cornell.edu/abusch/Abusch-Pragmatic-Presupposition.pdf.

Bourgeois-Gironde, Sacha and Giraud Raphael. 2009. "Framing Effects as Violations of Extensionality". *Theory and Decision*. 67 (4): 385 – 404.

Fălăuş, Anamaria. 2013 "Introduction: Alternatives in Semantics and Pragmatics". In Anamaria Fălăuş (ed.) *Alternatives in semantics. Palgrave Studies in Pragmatics, Language and Cognition*. Palgrave Macmillan. 1 – 35.

Grice, Herbert Paul. 1975. "Logic and Conversation". In Peter Cole and James L. Morgan (eds.) *Syntax and Semantics. vol. 3. Speech Acts*. New York: Academic Press. 41 – 58.

Ingram, Joanne. 2010. *Focus, polarity and framing effects*. PhD thesis. http://theses.gla.ac.uk/1446/.

Jaszczolt, Katarzyna M. 2010/2015. "Default Semantics" In Bernd Heine and Heiko Narrog (eds). *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 743 – 771.

Jaszczolt, Katarzyna M. 2014. "Defaults in Semantics and Pragmatics". In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition). https://plato.stanford.edu/archives/win2014/entries/defaults-semantics-pragmatics.

Jones, Martin K. 2007. "A Gricean Analysis of Understanding in Economic Experiments". *Journal of Economic Methodology*. vol. 14. n. 2. 167 – 185.

Kahneman, Daniel and Tversky Amon. 1979. "Prospect Theory: An Analysis of Decision under Risk". *Econometrica*, 47(2): 263 – 291.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.

Kühberger, Anton. 1998. "The Influence of Framing on Risky Decisions: A Meta-Analysis". *Organizational Behavior and Human Decision Processes*. 75(1): 23 – 55.

Larrick, Richard P., Smith Edward Elmer and Yates Frank J. 1992. "Reflecting on the Reflection Effect: Disrupting the Effects of Framing Through Thought". Paper presented at the meetings of the society for Judgment and Decision Making, St. Louis, November 1992.

Levin, Irwin P., Schneider Sandra L. and Gaeth Gary J. 1998. "All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects". *Organizational Behavior and Human Decision Process*. vol. 76. n. 2. November. 149–188.

Levinson, Stephen C. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*, Cambridge, (MA): MIT Press.

Maheswaran, Durairaj and Meyers-Levy Joan. 1990. "The Influence of Message Framing and Issue Involvement". *Journal of Marketing Research*. vol. 27. n. 3. August. 361–367.

Marteau, Theresa. M. 1989. "Framing of Information: Its influence upon Decisions of Doctors and Patients". *British Journal of Social Psychology*. v. 28. n. 1: 89–94.

Mayer, Rolf. 1992. "To Win and Lose: Linguistic Aspects of Prospect Theory". *Language and Cognitive Processes*. vol. 7. n. 1: 23–66.

McKenzie, Craig R.M. and Nelson Jonathan D. 2003. "What a Speaker's Choice of Frame Reveals: Reference Points, Frame Selection, and Framing Effects". *Psychonomic Bulletin & Review*. vol. 10. n. 3. 596–602.

Millar, Murray G. and Millar Karen U. 2000. "Promoting Safe Driving Behaviors: The Influence of Message Framing and Issue Involvement". *Journal of Applied Social Psychology*. vol. 30. n. 4. April. 853–866.

Moscati, Ivan. 2012. "Intension, Extension, and the Model of Belief and Knowledge in Economics". *Erasmus Journal for Philosophy and Economics*. v. 5. n. 2. Autumn. 1–26.

Moxey, Linda. 2011. "Mechanisms Underlying Linguistic Framing Effects". In Gideon Keren (ed.) *Perspectives on Framing*. New York: Psychology Press. Tylor & Francis Group. 119–134.

Nardo, Michela. 2003. "The Quantification of Qualitative Survey Data: A Critical Assessment". *Journal of Economic Surveys*. vol. 17. n. 5. 645–668.

Russo, Edward, J., Medvec Victoria Husted and Meloy Margaret G. 1996. "The Distortion of Information during Decisions". *Organizational Behavior and Human Decision Process*. vol. 66. n. 1. April. 102–110.

Sbisà, Marina. 1999. "Presupposition, Implicature and Context in Text Understanding". *Proceedings of the Second International and Interdisciplinary Conference on Modeling and Using Context*. London: Springer-Verlag. 324–338.

Schoorman, David F., Mayer Roger C., Douglas Christina A. and Hetrick Christopher T. 1994. "Escalation of Commitment and the Framing Effect: An Empirical Investigation". *Journal of Applied Social Psychology*. vol. 24. n. 6. March. 509–528.

Searle, John. 1978. "Literal Meaning". *Erkenntnis*. 13. 207–224.

Sher, Shlomi and McKenzie Craig R.M. 2006. "Information Leakage from Logically Equivalent Frames". *Cognition*. 101. 467–494.

Simons, Mandy. 2003. "Presupposition and Accommodation: Understanding the Stalnakerian Picture". *Philosophical Studies*. 112: 251–278.

Simons, Mandy. 2013. "Presupposing". In Marina Sbisà and Ken Turner (eds), *Pragmatics of Speech Actions*. Berlin: De Gruyter Mouton, 143–172.

Stalnaker, Robert. 1974/1999. "Pragmatic presuppositions". In: Stalnaker Robert, *Context and Content*. Oxford: Oxford University Press. 47–62.

Stalnaker, Robert. 2002. "Common Ground". *Linguistic and Philosophy*. 25: 701–721.

Schwarz, Norbert. 1996. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*. Hillsdale. NJ: Erlbaum.
Taylor, S. E. 1991. "Asymmetrical Effects of Positive and Negative Events: The Mobilization-Minimization Hypothesis". *Psychological Bulletin*. 110. 67 – 85.
Tversky, Amon and Kahneman Daniel. 1981. "The Framing of Decisions and the Psychology of Choice". *Science*. vol. 211. n. 4481. January. 30. 453 – 458.
Tversky, Amon and Kahneman Daniel. 1991. "Loss Aversion in Riskless Choice: A Reference-dependent Model". *Quarterly Journal of Economics*. vol. 107. November. 1039 – 1061.

Rasmus Gahrn-Andersen

Transcending the Situation: On the Context-dependence of Practice-based Cognition

Abstract: Socio-material practices entail what Per Linell (2009) terms 'situation-transcendence' in that they allow agents to engage in purposeful activity that exceeds the spatiotemporal constraints of a given situation. Radical perspectives in cognitive science offer insights into the cognitive dynamics involved: while those supportive of Distributed Cognition (DCog) analyze how cognition is distributed synchronously across situations, proponents of 4E Cognition show that cognition extends in a diachronic manner. Despite their differences, both perspectives acknowledge that agents must be sensitive to contextual information in order to accomplish cognitive tasks. In fact, both analyze the functionality of socio-cultural resources (e.g. computers, notebooks, instruments etc.) in relation to particular contexts and tasks, and focus on how such ressources give rise to the cognitive resources required for different kinds of cognitive activity. Having described key positions in the literature, the paper presents a case study on the practice of leakage detection in heating pipes as this is done by professionals in a Danish utility company. Accordingly, the paper explores situation-transcendence and, more specifically, how contextual constraints connect with and emerge from the synchronic and diachronic processes involved as professionals engage in data analysis, internal and external coordination and on-site exploration.

1 Introduction

Human cognition unfolds in a cultural-cognitive ecosystem consisting of a myriad of different socio-material practices. According to Edwin Hutchins (2010), we should bear in mind the following: Practices are essential to human living in that they shape our individual and collective ways of making sense of our surroundings.[1]

1 It is because of our embeddedness in practices that we can perceive something *as something*—such as, for instance, a line of people as a queue (Hutchins 2014), one thing as a hammer (Hei-

Rasmus Gahrn-Andersen, University of Southern Denmark (rga@sdu.dk)

https://doi.org/10.1515/9783110702286-012

When studying these practices, we ought to focus on "cognitive phenomena in context" (Hutchins 2010: 795). Accordingly, it is imperative that we explore practical activity in real-life settings and not just stick to theory.[2] In this connection, it is important to abstain from focusing exclusively on either macro-level or micro-level descriptions (cf. Latour 1996a). One must carefully avoid appealing to the functional characteristics of particular practices or social systems as do, for instance, Émile Durkheim (1982) and Niklas Luhmann (1995). Macro-views should be avoided because they are typically less concerned with the skills and competencies that agents bring to the fore. Thus, proponents of these views are reluctant to explore the context-sensitive dynamics involved as agents exploit different technologies to perform given tasks. But we must also resist the temptation of digging in too deep by focusing on the micro and, consequently, limit our focus to interactions in a particular context (e. g. Steffensen *et al.* 2016) or situation (e. g. Goodwin 2000). In so doing, one risks losing sight of the normative constraints that influence interactional outcomes. We should rather focus on the 'in-between' that connects micro- and macro levels (cf. Latour 1996a).

Exploring this intersection entails focusing on how practice-embedded agents allow different situations (with their localized and unique contexts) to interrelate. Bruno Latour (1996a; 1996b; 1999) was amongst the first to conceptualize this in-between. According to him, social phenomena are irreducible to a situation or its spatiotemporal constraints, meaning that human socio-material engagements are always "framed by other actants dispersed in space and time." Latour emphasizes this by arguing that situations are always shaped by events that occur in different contexts: "At time *t*, I find myself in contact with beings who have acted at *t-1*, and I fold the situations together so that I myself will act under another form at *t+1*. In situation *s*, I find myself attached to situations *s-1*, and I act such that downstream situations *s+1* come to be associated with mine" (Latour 1996: 239).

---

degger 2010), another thing as a tomato (Noë 2006) etc. Practices give rise to socially-informed ways of understanding that saturate our engagements with the world: not only do they allow us to identify things that have social significance; they also enable us to act-perceive, or *perc-act* (Berthoz 2012) in accordance with others as we, for instance, set out to solve a problem (Steffensen 2013), ask someone to hand us a tool (Wittgenstein 2009) or perform a complex task such as landing an airplane (Hutchins 1995b).

2 Following the Merriam-Webster Dictionary, I take 'context' to denote "the interrelated conditions in which something exists or occurs" (https://www.merriam-webster.com/dictionary/context accessed May 19, 2018). Consequently, being serious about context "means finding oneself in the thick of the complexities of particular situations at particular times with particular individuals" (Nardi 1996: 35).

Depending on the situation, the shaping can be more or less structured. For example, in professional practices, we are not just witnessing the 'ongoing' or loose-coupling of plain everyday activities such as greeting a random person at a bus stop. Rather, these practices are structured in the sense that cognition forms "part of a larger set of 'equipment', enhanced and fostered by human co-operation" (Secchi and Cowley 2018: 1). Practice-informed cognitive activity occurs not just in *a* situation, but, crucially, *across* situations. In the terms used by linguist Per Linell (2009), practices involve *situation-transcendence*,[3] which is "linked to habituality, routinization, conventionalization and institutionalization of human practices, that is, our tendencies to do things approximately in the same ways as have done before, or seen others do, in similar situations" (Linell 2009: 50).

Accordingly, the purpose of this paper is to explore how practices afford situation-transcendence as agents solve tasks while drawing on cognitive skills, environmental resources and contextual contingencies. The paper is structured as follows: first, I turn to radical embodied cognitive science and show how proponents of Distributed Cognition (DCog) and 4E Cognition account for the situation-transcendence and context-dependence of human cognition. Whereas the DCog-perspective is well-known for its focus on synchronous activity involving real-time interaction between agents, the 4E-perspective offers other insights by underlining the importance of individual skills and the diachrony of cognitive phenomena. Second, I present a case study on situation-transcendence in a Danish utility company. The study focuses on the practice of leakage detection and how situation-transcendence emerges as professionals engage in distributed and extended cognitive processes that are in part shaped by contextual factors.

2 Mixing DCog and 4E Cognition

Traditional approaches to philosophy of mind and cognitive science are based on the assumption that we can understand cognition by focusing on the brain (or mind) of subjects in isolation from worldly structures (cf. Noë 2009). In philosophy, this is exemplified by the well-known 'brain in a vat'- thesis. 'Radical' ap-

3 A similar insight is found elsewhere in linguistics. For example, linguistic anthropologist Nick Enfield (2014) suggests that linguistic activity occurs across six different time-scales: 1) the microgenetic, 2) the ontogenetic, 3) the phylogenetic, 4) the enchronic, 5) the diachronic and, finally, 6) the synchronic. In this view, human sociality is multi-scalar in that it involves a range of different temporalities that exert their influence on the present activities (see also Steffensen and Pedersen 2014).

proaches to cognition are different in that they explore agent-environment relations, thereby presuming that cognition involves agent-environment dynamics that exceed the boundaries of the skull (Chemero 2009). They are thus typically characterized by a commitment to *contextualism* in the sense that sense-making practices are assumed to be "situated and contexted" (Linell 2010: 11). But when it comes to exploring the situation-transcendence of cognitive activity, DCog and 4E Cognition bring about different insights.[4]

2.1 DCog

DCog came out of California in the 1980s. Its supporters rightly claim that cognition does not simply happen inside people's heads but is distributed across media, people, artefacts and tools. For the last 30 years, DCog research has systematically investigated distributed cognitive processes in different, often highly specialized work-environments as professionals set out to solve complex tasks. It thus focuses on how professionals coordinate their actions while drawing on the externalized representations offered by different technologies: "The conduct of the [cognitive] activity proceeds by the operation of functional systems that bring representational media into coordination with one another. The representational media may be inside as well as outside the individuals involved. These functional systems propagate representational state across the media" (Hutchins 1995a: 372).

For example, in a famous study on a crew navigating a U.S. Navy vessel, Hutchins (1995a) accounts for how various cognitive tasks are performed by sailors while using different tools and instruments (e.g. fathometers, the pelorus, charts, walkie-talkies etc.). Other studies of DCog have explored cognition in different practices such as, for instance, piloting an airplane (Hutchins 1995b), air-traffic control (Halverson 1995), engineering teams (Perry 2013) or crime scene investigation (Baber 2010)—just to name a few. One obvious strength of DCog is its focus on how different artefacts and technologies are put to use by professionals in highly specialized practices as they attempt to solve specific tasks or problems on hand.

---

4 The radical nature of these approaches has been questioned. For instance, Graham Button argues that DCog is far from as radical as it initially suggests in that its proponents think that there are internal neurological processes at play as well. DCog should thus not be seen as "a corrective to traditional cognitive science" but as an attempt "to make the socio-cultural world a further topic for cognitive science" (Button 2008: 102). Also, Daniel Hutto and Erik Myin (2013) show that many enactivist approaches make use of 'internalist language' characteristic of traditional cognitive science.

Interestingly, DCog also explores the context-dependence of cognition. For instance, in his study of the practices onboard USS Palau, Hutchins provides the example of avoiding a collision with another ship. He describes the crew's coordination as involving synchronous situation-transcending activities in the sense that the distributed cognitive system extends, in real-time, from the pilothouse on the bridge to the flight deck as the crew try to avoid the collision (Hutchins 1995a: 4–5). Specifically, the navigator instructs the keeper of the deck log to locate a foghorn so that the crew can alert the other ships' crew. In this connection, the keeper of the deck log needs to descend to the flight deck in order to retrieve the horn. In doing so, he continues to communicate with the bridge by means of his walkie-talkie (Hutchins 1995a: 5). Thus, the walkie-talkie facilitates cognition in that it enables the crew to coordinate their activities in a synchronous manner despite the difference in their situatedness that spans across different locations on the ship.

Despite also acknowledging that cognition "may be distributed through time in such a way that the products of earlier events can transform the nature of later events", DCog researchers are usually more interested in synchronic activities that happen immediately across situations and contexts (cf. Nardi 1996: 42). Thus, as Hyundeuk Cheon, summarizes: "The operation of a distributed-cognitive system is parallel in that multiple people and artifacts work simultaneously" (Cheon 2014: 24).

In focusing on the simultaneity, or what Cheon terms "parallelism", of the distributed cognitive processes of a given cognitive system, there is a tendency in DCog research to neglect the fact that socio-cognitive activity is also diachronically linked and, thus, historically shaped across situations.[5]

5 The criticism reaches beyond DCog. For instance, in her micro-sociological account, Karin Knorr Cetina (2009) explores how modern technology such as computers, TVs and the internet have brought a fundamental change to localized interactions. In cases with technological mediation and presence, social interactions are irreducible to the embodied factors of Goffman's (1983) 'natural situation'. Cetina contrasts Goffman's natural interactions with the 'synthetic', arguing that the latter "also implies encounters in time and the temporal extension and specification of situations" (Cetina 2009: 62). Despite emphasizing the temporal dimension, Cetina focuses predominantly on 'real-time contexts' where the synthetic situation "is a composite of information bits that may arise from many areas around the world and feature the most diverse and fragmented content" (69). In the view presented here, the interesting thing is also how different real-time contexts are connected, thus reflecting the complexity (or compounded nature) of human socio-material practices and their situation-transcendence.

2.2 4E Cognition

This brings me to 4E Cognition—a label that was first introduced by Mark Rowlands (2010) to denote those approaches that criticize cognitive internalism by arguing that cognition is extended, embodied, embedded, and enacted. These theories are in basic accord with the "socio-cultural dimension" of DCog (cf. Button above) in that they build on cognitive externalism. On this view, cognitive activity scaffolds—or, extends—beyond the brain of the cognizer and into the environment (cf. Colombetti and Krueger 2015). As I argue below, 4E approaches have the potential for supplementing DCog in a way that makes the combination of the two highly relevant for exploring how practices unfold as situation-transcending activity.

However, it is first worth mentioning that in contrast to traditional DCog-approaches, which tend to focus on a particular distributed cognitive system in its entirety, 4E theories explore the individual agents involved and, more specifically, subject-relative factors such as experience and embodiment. For instance, proponents of the so-called *embodiment thesis* recognize that cognizers are always physically situated individuals.[6] Following Daniel Hutto and Erik Myin, it is a common assumption that "no matter how the empirical questions are answered, mentality [...] is in all cases concretely constituted by, and thus literally consists in, the extensive ways in which organisms interact with their environments" (Hutto and Myin 2013: 7).

Here, there is a clear link to enactivist thinking and the idea promoted by Francisco Varela (1987) that cognitive activity should be considered as analogous to "laying down a path in walking." We should therefore not consider cognition as pre-given mental content but rather as an activity enacted by agents as they engage with the world. Strongly embodied cognition is thus bound to the specific context in which the agent is situated, meaning that cognition is a strongly synchronous phenomenon. This is also reflected in the notion of 'participatory sense-making' which takes instances of 'corridor dancing' as a prime example of how basic coordinative behavior plays out (see, De Jaegher and Di Paolo 2007; for a criticism, see Cowley and Gahrn-Andersen 2017).

But as the label suggests, proponents of 4E Cognition do not take cognition to be irreducible to its embodied and enactive aspects as do, for instance, enactivists. This brings me to another useful aspect of the 4E perspective: the idea

6 Interestingly, Sterelny (2011) differs in that he oscillates between applying 'scaffolding' on a population level and in specific contexts or in relation to the evolution of particular technologies such as, for instance, tools or bodily mechanisms such as the human digestive system.

that cognition is extended.[7] Clark and Chalmers (1998) introduce their Extended Mind Hypothesis by arguing that human agents exploit external artefacts (e. g. notebooks, blackboards, computers etc.) as tools for extending their cognitive activities. Thus, cognizers do not merely rely on their intrinsic cognitive powers, but draw on various media containing externalized representations (cf. Heersmink 2017). This seems to be much in line with DCog. So, how do the two differ and to what extent can insights related to the extendedness of cognition supplement distributed approaches and, further, provide relevant insights into the situation-transcendence of human cognition?

On the view presented here, extended cognition offers at least one valuable insight. This is evident from Clark and Chalmers' (1998) example with the fictitious Alzheimer's' patient, Otto. Because of his illness, Otto has a limited internal memory capacity and relies on external artefacts (e. g. a notebook) to aid his memory. Thus, "he relies on information in the environment to help structure his life. Otto carries a notebook around with him everywhere he goes. When he learns new information, he writes it down. When he needs some old information, he looks it up. For Otto, his notebook plays the role usually played by a biological memory. Today, Otto hears about the exhibition at the Museum of Modern, and decides to go see it. He consults the notebook, which says that the museum is on 53rd Street, so he walks to 53rd Street and goes into the museum" (Clark and Chalmers 1998: 12–13).

Whereas some of those in favor of DCog tend to overemphasize the importance of particular technologies and their 'access characteristics' (Perkins 1993), Clark and Chalmers underline the necessity of (inner) mental correlates and, tacitly, the context-dependence of cognition. They acknowledge that cognitive states (or parts hereof) must necessarily be extended, thus presuming a stable element in Otto's phenomenology: he must have what, in lack of better words, one could call a *partial belief* in the museum in order to be able to make proper use of his notebook. Accordingly, "*beliefs* can be constituted partly

7 There are three waves of the extended cognition-thesis: defenders of the first-wave focus on the cognitive system of individuals and how inner and outer elements are functionally similar (cf. Kirchhoff 2015). Proponents of the second-wave emphasize, among other things, the importance of integrating cognitive processes into larger wholes (see Menary 2007). Lastly, the third-wave "dissolves individuals into peculiar loci of coordination and coalescence among multiple structured media [and practices]" (quote by Sutton in Kirchhoff 2015). Michaelian and Sutton summarize: "Clark saw distributed cognition as a natural extension of the point made in much connectionist literature that order and systematicity in human cognition and action can derive in part from the stability of our environments, rather than as a direct product or reflection of exhaustively-specified internal recipes. Cognition might thus be multiply distributed, both within neural networks and across bodies, artifacts, and social groups" (Michaelian and Sutton 2013: 6).

by features of the environment, when those features play the right sort of role in driving cognitive processes" (Clark and Chalmers 1998: 12).

The example with Otto underlines that extended cognition need not only occur in a particular, natural situation (to paraphrase Goffman 1983); it can also facilitate diachronic situation-transcendence. The context-dependence is evident from the fact that Clark and Chalmers tacitly presuppose that Otto uses his notebook every time he forgets the address; even as he is on his way to the museum. This happens across situations and contexts. Otto's pre-existing belief in the museum, its relevance etc. enables him to walk past various locations (e.g. his own house, the street he lives on etc.) on his way to the museum without ever losing sight of his destination.

Having shown that DCog and 4E perspectives build on important, but nevertheless different assumptions concerning key characteristics of human cognition,[8] I will now present empirical findings that show how practice-based activity relies not only on context-dependence but also on situation-transcendence, while exploring its synchronous and diachronic aspects.

3 Situation-transcending Cognition: the Example of Leakage Detection

3.1 Background of the study

The data was collected as part of an ethnographic study exploring the extent to which the use of drones in leakage detection affects routines and decision-making processes in a Danish utility company.[9] Leaks in heat pipes are a common problem for providers of district heating. Every year, the Danish utility sector pumps millions of liters of make-up water into what were designed to be closed networks of pipes connecting their boilers with radiators across the nation. Ex-

8 As an anonymous reviewer points out, one major difference between the two approaches is that while supporters of Radical Enactive (or Embodied) Cognition are critical of mental representations (cf. Hutto and Myin 2013: xii), proponents of DCog have no issue in retorting to mental content when explaining cognition (take, for instance, Hutchins' (2005) notion of *cognitive projection*).

9 This exploratory ethnographic study was conducted in March and April 2018. Using a methodological blend of *participant observation* (Emerson *et al.* 2007) and informal *semi-structured interviews* (Edwards and Holland 2013) I collected different kinds of data including fieldnotes, video-recordings and photographs. The overall purpose of the study was to establish the role *TeraPlan* plays in the utility company's daily maintenance operations.

tending for more than 60.000 kilometers, the network of heating pipes supplies heat to 1,7 million households and 3,5 million people.[10] Leaks often occur due to erosion and general wear and tear. But they are difficult to spot because the pipes are located below the ground's surface. Also, far from all pipes have an in-built warning system that signals the occurrence of a damage. With the purpose of boosting efficiency, several Danish utility companies have begun collaborating with companies specializing in using drones for thermographic surveillance. They are thus able to spot damages that are invisible to the naked eye because the thermographic camera is able to detect unusual heat radiation coming from the ground's surface.

In the current study, the utility company arranges for their contractor to overfly a particular area and photograph it accordingly. Subsequently, the contractor makes the imagery available to the utility company in *TeraPlan*; an internet-based software that has *Google Maps* as its basis. It is then up to the utility company's employees, formally employed as 'coordinators', to access the data and decide if action is needed. Having established the overall frame of the study, I will now turn to a specific practice of leakage detection that not only involves the two kinds of situation-transcending activity introduced above but also serves to show how context-relative factors influence sense-making processes.

3.2 The office

The primary situation is the main office where the utility company's coordinators spend most of their time. They access *TeraPlan* from their computers in order to gain an overview of possible leaks in a given area of the city. This typically happens weeks (and sometimes even months) after the area has been overflown. The reason for this is that the contractor's IT-specialists first have to manually update the software by screening the thermographic images for areas with increased heat radiation and map these accordingly. The procedure is as follows: they mark the most evident leaks with an A, less evident leaks are given a B and potential leaks that should be kept under observation are marked C. Lastly, E is used to designate special cases that initially show signs of leakages but which the IT-specialists have ruled out. Depending on the classification, the coordinator in the utility company will initiate different activities. For example, in areas marked A, he will simply check the thermographic images and compare them

10 http://www.danskfjernvarme.dk/presse/fakta-om-fjernvarme information retrieved on May 19, 2018.

with the *Geographic Information System*, or *GIS*, that depicts the utility company's network of tubes and the technical specifications of that area of the network (see figure 1). He does so in order to get a proximation of the location of the leak, the length of the pipe, the kind of pipe etc. Subsequently, he will arrange for a contractor specialized in digging to begin the works in the designated area. All of this takes place in the primary situation given that the coordinator can do everything on his computer in the office. There are situations, however, where areas in *TeraPlan* are marked with Bs and Cs. Here, the practice of leakage detection entails additional steps and a secondary situation. Before getting to this, I wish to clarify the extent to which the primary situation and the context-dependence of the cognition involved allows for situation-transcendence.

Figure 1: The coordinator situated in front of his computer using multiple screens to compare data from *TeraPlan* (right screen) with the technical information from the *GIS* (left screen).

Distributed cognition: *TeraPlan* is designed to serve a distributed cognitive system that involves basic coordination between, on the one hand, the contractor's drone-operators and IT-specialists and, on the other, the coordinators in the utility company. Simply put, the software supplies information that the coordinators can access. The contractor is responsible for making sure that the drone

overflies a designated area and that the software provides an extensive overview of suspected leaks in the designated parts of the city. Also, the contractor might decide to update the software with new features or by adding a set of newly discovered leakages. Thus, it could seem that the coordination happens in a somewhat unidirectional way; from the contractor to the utility company. Nevertheless, if a coordinator is in doubt about something, he will simply call or send an email to the contractor. He would do so, for example, if the drone had overflown a smaller area than agreed upon or if he needed to extend (or limit) a colleague's user-rights in *TeraPlan*.

But the organization involves other contractors. For instance, when facing an A, a coordinator will get in touch with a digging company using a different software, *Entrepriseportalen* ('*The Work Portal*'). Once an order has been issued, the digging company receives a notification via email. Also, since there are usually several people working with *TeraPlan* at the same time, the coordinators use its features to coordinate their actions. For example, when a coordinator considers a given leak, he will immediately change its designation (i. e. from A, B or C) to D. He does so to signify 'work-in progress'. Once the leak has been repaired, it will be reclassified as a F.

All of this occurs in real-time contexts and is enabled as well as constrained by computer software. In addition, it so happens that the coordinators discuss the data amongst themselves, thus providing a prime example of what Cetina calls an *electronically projected situation*. This kind of situation, she argues, "reaches far beyond what would ordinarily be visible in a physical setting; not only does it include many layers and windows providing geopolitical and epistemic depth and internal contextualization, but it also stitches together an analytically constituted world made up of 'everything' potentially relevant to the interaction" (Cetina 2009: 65 – 66).

The practice of leakage detection involves a distribution of cognitive processes and coordination internally in the utility company as well as externally. Evidently, it is highly context-dependent as, for instance, changes in classifications in *TeraPlan* afford changes in the coordinators' sense-making activities. I will now turn to some of the extended, embodied, embedded, and enacted factors involved as a coordinator makes sense of the data marked either as a B or a C.

Making sense of the data, structuring the context: Since the thermographic image is an optional layer in *TeraPlan*, the coordinator first has to activate it. Subsequently, it needs to be aligned with the satellite image from *Google Maps* (see figure 2) since it would be more or less randomly placed on the map. However, the coordinator needs to know exactly where the suspected leak is located because the thermographic image contains no meta-data such as street name, house number etc.. For this reason, he engages in what David Kirsh

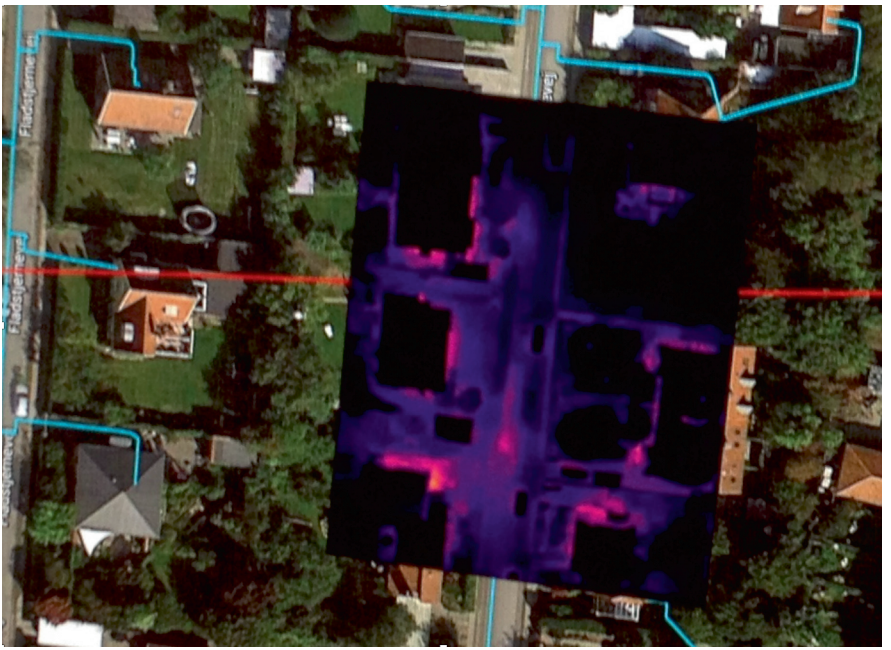

Figure 2: Close-up of *TeraPlan*. Notice the customized satellite image from *Google Maps*. The blue lines signify the network of heating pipes while the red line shows the drone's flightpath. Further, the thermographic image in the center depicts different thermal radiations with the potential leak in the middle of the road. Missing from the screenshot are the markers (i. e. the As, Bs, Cs, Ds, Es and Fs) that indicate suspected leaks, works in progress and completed works. I would like to thank Drone Systems for allowing me to use this screenshot.

and Paul Maglio (1994) term *epistemic actions*, which are "performed to uncover information that is hidden or hard to compute mentally" (Kirsh and Maglio 1994: 513). Like matching blocks in a game of Tetris—but without the same time pressure—the coordinator carefully adjusts the thermographic layer to the satellite image. In so doing, he does not only depend on the actual context, he also affects both the context and his "own computational state[s]" (Kirsh and Maglio 1994: 514).[11] Specifically, he manually rotates the layer until it 'fits' the background. In so doing, he projects his cognition onto the computer screen while

11 The idea that action shapes contexts is commonly held by proponents of *Activity Theory* (e. g. Leontiev 1978). Proponents of this position rightly assume that a context "is constituted through the enactment of an activity involving people and artifacts" (Nardi 1996: 38). In other words, it makes little sense to think that a context remains unaffected by the actions that individuals perform.

fiddling with the layer.[12] Although his actions occur in the environment (rather than inside his head), the coordinator nevertheless correlates them with some standard (or intrinsic value) in the sense that he wants the two images to be 'perfectly aligned'.

Once this has happened, he will try to estimate the precise location of the leak by drawing on information from the *GIS*. He wants to get a sense of, for instance, if there are any valves nearby, if maintenance has been done on a particular part of the network etc. In this connection, he enacts a '[virtual] trajectory' (cf. Hutchins 2014) as he links pipe-specifications from the *GIS* with *TeraPlan*'s thermographic layer. In so doing, he imagines the location of the pipes and which of their components might be affected by the leak. As figure 1 shows, this involves comparing information on two computer screens. This part of the leakage-detection is vital because it is a way for the coordinator to prepare himself before entering the second phase, which entails a diachronic situation-transcending activity in that it extends into a new situation.

3.3 On-site

The coordinator drives to the location of the suspected leak depicted in the thermographic image which, consequently, becomes the secondary situation involved in the leakage detection. The interrelatedness of the two situations underlines Linell's point that situation-transcendence entails that "contexts are dynamic; they become relevant and emerge in and through interaction" (Linell 2009b: 19). Indeed, situation 2 is highly dependent on the actions happening in situation 1 including that the coordinator agrees with the contractor's classifications in *TeraPlan*. The coordinator typically visits the location when he goes either to or from work. But the two situations typically occur days apart. Regardless, they are connected in a diachronic manner for the simple reason that they belong to different temporal domains with their unique contexts. Their cognitive 'connector' can be cashed out in 4E-terms: The coordinator has a partial belief in a leakage (similar to the one Otto has in the museum) which he intends to test by saturating his experiences or using what is known as *interactivity* (cf. Harvey *et*

12 In fact, Clark and Chalmers (1998) evoke a similar example with a person sitting in front of a computer screen who has the possibility to either "physically rotate the image on the screen, by pressing a rotate button, or to mentally rotate the image." And we "can also suppose, not unrealistically, that some speed advantage accrues to the physical rotation operation." (Clark and Chalmers 1998: 7). In line with Kirsh and Maglio's point, the coordinator does not bother spending time on mentally aligning the two images.

al. 2016). His partial belief motivates him to drive to the site which is sometimes located several kilometers away from the office. He typically uses the GPS in his smartphone to find the way. Once arriving on the site, he will access the *GIS* (also on his smartphone) and take out a print-out of the thermographic image from *TeraPlan*. He then simply compares the data he accessed at the office with what he sees at the actual location.

Figure 3: The site of the suspected leak. I would like to thank Søren Ernst (HOFOR A/S) for allowing me to use this photo.

Depending on the context, the site can afford different useful insights to the coordinator. For instance, if it is located in a residential neighborhood, the coordinator may choose to ask residents if they have experienced anything that could indicate a leakage including 1) if they have seen green water either outside on the street, on their premises or in their basements,[13] or 2) if there are places on the street that do not freeze in winter. It also happens that the site is not a

13 NB. The utility company colors their make-up water so that residents will notice if water used in the district heating accidentally mixes with drinking water.

residential area but a heavily trafficked road. Nevertheless, the coordinator still relies on cues offered by the situation. He typically resorts to using his *professional vision* (Goodwin 1994) that entails a degree of diachrony in that it draws heavily on his competencies, skills and professional experience. This kind of vision enables skilled professionals to directly pick up available contextual information. The information is typically non-salient or hidden to non-professionals and is also context-specific. For example, if there has been snow or rainfall a few hours earlier, the coordinator knows that a leakage might be visible to the naked eye in that it would leave either a wet spot in the case of snow or, in the case of rain, a dry spot (see figure 3).

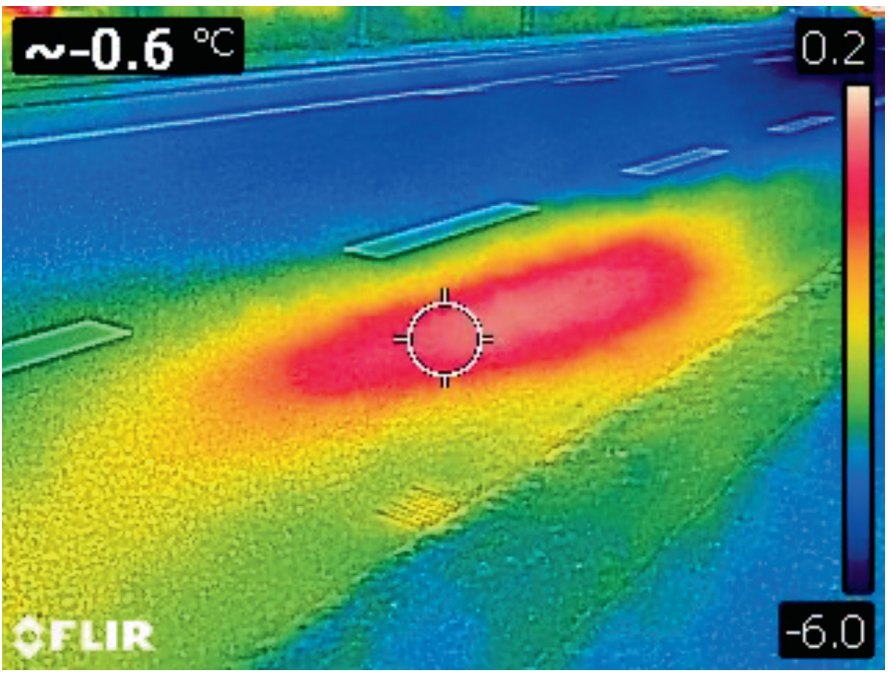

Figure 4: Picture from the coordinator's thermographic camera showing a suspicious increase of heat radiation. I would like to thank FLIR Systems for allowing me to include this image.

Regardless of the contextual evidence, the coordinator will verify his suspicion by taking photos with his thermographic camera (see figure 4) which has greater sensitivity than either his professional vision or that of the drone. Having reached saturation, he will assign the diggers. He can do this either straight away by means of his smartphone or once he is back in the office.

4 Concluding Remarks

The case study exemplifies how situation-transcendence occurs 'in the wild' in the context of a utility company. Socio-material practices entail situation-transcendence in two different but mutually supportive senses. In its synchronous sense, situation-transcendence involves a real-time distribution of cognitive resources across media and agents. Diachronic transcendence, on the other hand, reflects that cognition has a historical basis that enables practice-specific cognitive phenomena including the skills and competencies that a cognizer brings to the fore to solve tasks in particular situations. In addition, the study also shows that much can be gained from considering cognition in relation to socio-material practices. In this connection, the study contributes to research conducted under the 4E-label in at least two ways.

First, it eases the tension between enactive and extended perspectives by underlining that extended cognitive processes need not solely be cashed out in 'substrate-independent' terms as Di Paolo (2009) has argued (cf. Clark & Kiverstein 2009). In the case of the utility company, the extended processes themselves come to function as substrates in that the normative regulation of different cognitive causal couplings can be traced to the coordinator's partial beliefs. Partial beliefs play an indispensable role since they functionally connect the coordinator's sense-making of *TeraPlan*'s externalized representations (e. g. the thermographic images, the categorizations) with the sense-making that occurs on the site of the suspected leaks. These beliefs thus come to be normatively constitutive of the relations between the different cognitive activities that comprise the practice of leakage-detection.

Second, it questions the body-centrism that is central to the embodied cognition approach (Kersten *et al.* 2017). By assuming that the body is 'intrinsically special' (2395), proponents of this approach tend to focus on highly situated cognitive activity (cf. Rupert 2010; see, for instance, Rowland's (2010: 198 – 200) three examples of intentionality as 'travelling through'). When it comes to understanding complex socio-cognitive phenomena, however, one should also consider the different interdependencies that enable their occurrence—and, equally important, reoccurrence. For example, the coordinator's *professional vision* and, thus, his sensitivity to environmental cues are shaped by past encounters with leakages while his routine of aligning images in *TeraPlan* is habitually construed by his knowledge of the software. By taking a practice-based view, the study shows that although the body plays a crucial role, the agent's situated embodied cognition remains heavily influenced by past interactions. For this reason, a

trans-situational approach informs situational approaches when it comes to exploring the significance of human practical activity.

Acknowledgements

This ethnographic study was conducted as an independent part of the project *Organizational Barriers to the Implementation of Drones* founded by the University of Southern Denmark's strategic *Lighthouse*-initiative. I am grateful to principal investigator Stephen J. Cowley for his valuable comments on an earlier version of this paper. I would also like to thank Bo Jensen Møller (HOFOR A/S) for making the study possible.

References

Baber, Chris. 2010. "Distributed Cognition at the Crime Scene". *AI and Society* 25(4). 423–432.

Berthoz, Alain. 2012. *Simplexity: Simplifying Principles for a Complex World*. New Haven and London: Yale University Press.

Button, Graham. 2008. "Against 'Distributed Cognition'". *Theory, Culture & Society* 25(2). 87–104.

Cetina, Karin K. 2009. "The Synthetic Situation: Interactionism for a Global World". *Symbolic Interaction* 32(1). 61–87.

Chemero, Anthony. 2009. *Radical Embodied Cognitive Science*. Cambridge, (MA): MIT Press.

Cheon, Hyundeuk. 2014. "Distributed Cognition in Scientific Contexts". *Journal for General Philosophy of Science* 45(1). 23–33.

Clark, Andy and Chalmers David. 1998. "The Extended Mind". *Analysis* 58(1). 7–19.

Clark, Andy and Kiverstein Julian. 2009. "Introduction: Mind Embodied, Embedded, Enacted: One Church or Many?". *Topoi* 28. 1–7.

Colombetti, Giovanna and Krueger Joel. 2015. "Scaffoldings of the affective mind". *Philosophical Psychology* 28(8). 1157–1176.

Cowley, Stephen J. and Gahrn-Andersen Rasmus. 2015. "Deflating Autonomy: Human Interactivity in the Emerging Social World". *Intellectica* 63. 49–63.

De Jaegher, Hanne and Di Paolo Ezequiel A. 2007. "Participatory Sense-making: An Enactive Approach to Social Cognition". *Phenomenology and the Cognitive Sciences* 6(4). 485–507.

Di Paolo, Ezequiel A. 2009. "Extended Life". *Topoi* 28. 9–21.

Durkheim, Émile. 1982. *The Rules of Sociological Method*. New York: The Free Press.

Edwards, Rosalind and Holland, Janet. 2013. *What is qualitative interviewing*. London and New York: Bloomsbury.

Emerson, Robert M., Fretz Rachel I. and Shaw Linda L. 2007. "Participant Observation and Fieldnotes". In Paul Atkinson, Amanda Coffey, Sara Delamont, John Lofland and Lyn Lofland (eds.), *Handbook of Ethnography*, 352–368. London: SAGE.

Enfield, Nick J. 2014. *Natural Causes of Language: Frames, Biases, and Cultural Transmission*. Berlin: Language Science Press.
Goffman, Erwin. 1983. "The Interaction Order: American Sociological Association, 1982 Presidential Address". *American Sociological Review* 48(1). 1–17.
Goodwin, Charles. 1994. "Professional Vision". *American Anthropologist* 96. 606–633.
Goodwin, Charles. 2000. "Action and Embodiment Within Situated Human Interaction". *Journal of Pragmatics* 32(10). 1489–1522.
Halverson, Christine A. 1995. *Inside the Cognitive Workplace: New Technology and air Traffic Control*. Unpublished doctoral dissertation. San Diego: University of California.
Harvey, Matthew, Gahrn-Andersen Rasmus and Steffensen Sune. 2016. "Interactivity and Enaction in Human Cognition". *Constructivist Foundations* 11(2). 234–245.
Heersmink, Richard. 2017. "Extended Mind and Cognitive Enhancement: Moral Aspects of Cognitive Artifacts". *Phenomenology and the Cognitive Sciences* 16(1). 17–32.
Heidegger, Martin. 2010. Being and Time. New York: SUNY Press.
Hutchins, Erwin. 1995a. *Cognition in the Wild*. Cambridge, (MA): MIT Press.
Hutchins, Erwin. 1995b. "How a Cockpit Remembers Its Speeds". *Cognitive Science* 19(3). 265–288.
Hutchins, Erwin. 2005. "Material Anchors for Conceptual Blends". *Journal of Pragmatics* 37. 1555–1577.
Hutchins, Erwin. 2010. "Cognitive Ecology". *Topics in cognitive science* 2(4). 705–715.
Hutchins, Erwin. 2014. "The Cultural Ecosystem of Human Cognition". *Philosophical Psychology* 27(1). 34–49.
Hutto, Daniel and Myin Erik. 2013. *Radical Enactivism: Basic Minds Without Content*. Cambridge (MA): MIT Press.
Kersten, Luke, Dewhurst Joe and Deane George. 2017. "Resolving Two Tensions in 4E Cognition Using Wide Computationalism". *Proceedings of CogSci 2017*. 2395–2400.
Kirchhoff, Michael D. 2015. "Cognitive Assembly: Towards a Diachronic Conception of Composition". *Phenomenology and the cognitive sciences* 14(1). 33–53.
Kirsh, David and Maglio Paul. 1994. "On Distinguishing Epistemic from Pragmatic Action". *Cognitive Science* 18. 513–549.
Latour, Bruno. 1996a. "On Interobjectivity". *Mind, Culture, and Activity* 3(4). 228–245.
Latour, Bruno. 1996b. "On Actor-network Theory. A Few Clarifications plus More than a Few Complications". http://www.bruno-latour.fr/sites/default/files/P-67%20ACTOR-NET WORK.pdf.
Latour, Bruno. 1999. *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, (MA): Harvard University Press.
Leont'ev, Aleksei. 1978. *Activity, Consciousness, and Personality*. Englewood Cliffs, (NJ): Prentice-Hall.
Linell, Per. 2009. *Rethinking Language, Mind, and World Dialogically: Interactional and Contextual Theories of Human Sense-making*. Charlotte, (NC): Information Age Publishing.
Linell, Per. 2010. "With Respect to Bakhtin: Some Trends in Contemporary Dialogical Theories" In Karin Junefelt and Pia Nordin (eds.). *Proceedings from the Second International Interdisciplinary Conference on Perspectives and Limits of Dialogism in Mikhail Bakhtin*. 18–32. Stockholm: Stockholm University.
Luhmann, Niklas. 1995. *Social Systems*. Stanford: Stanford University Press.

Menary, Richard. 2007. *Cognitive Integration: Mind and Cognition Unbounded*. Basingstoke: Palgrave Macmillan.
Michaelian, Kourken and Sutton John, 2013. "Distributed Cognition and Memory Research: History and Future Directions". *Review of Philosophy and Psychology* 4, 1 – 24.
Nardi, Bonnie A., 1996. "Studying Context: A Comparison of Activity Theory, Situated Action Models and Distributed Cognition". In Bonnie Nardi (ed.), *Context and Consciousness: Activity Theory and Human-Computer Interaction* 35 – 52. Cambridge, (MA).: MIT Press.
Noë, Alva. 2006. "Experience of the World in Time". *Analysis* 66(1). 26 – 32.
Noë, Alva. 2009. *Out of Our Heads*. New York: Macmillan.
Perkins, David N. 1993. "Person-plus: A Distributed View of Thinking and Learning". In Gavriel Salomon (ed.) *Distributed Cognitions: Psychological and Educational Considerations* 88 – 110. Cambridge: Cambridge University Press.
Perry, Mark. 2013. "Socially Distributed Cognition in Loosely Coupled Systems". In Stephen J. Cowley and Frédéric Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*, 19 – 41. London: Springer.
Rowlands, Mark. 2010. *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*. Cambridge, (MA): MIT Press.
Rupert, Robert. 2010. *Cognitive systems and the extended mind*. Oxford: Oxford University Press.
Secchi, Davide and Cowley Stephen J. 2018. "Modeling Organizational Cognition: The Case of Impact Factor". *Journal of Artificial Societies and Social Simulation* 21(1)/13.
Steffensen, Sune V. 2013. "Human Interactivity: Problem-Solving, Solution-Probing and Verbal Patterns in the Wild". In Stephen J. Cowley and Frédéric Vallée-Tourangeau (eds.), *Cognition Beyond the Brain* 195 – 222. London: Springer.
Steffensen, Sune and Pedersen Sarah B. 2014. "Temporal Dynamics in Human Interaction". *Cybernetics and Human Knowing* 21(1 – 2). 80 – 97.
Steffensen, Sune V., Vallée-Tourangeau Frédéric and Vallée-Tourangeau Gaëlle, 2016. "Cognitive Events in a Problem-solving Task: A Qualitative Method for Investigating Interactivity in the 17 Animals Problem". *Journal of Cognitive Psychology* 28(1). 79 – 105.
Sterelny, Kim. 2010. "Minds: Extended or Scaffolded?". *Phenomenology and the Cognitive Sciences* 9(4). 465 – 481.
Sutton, John. 2006. "Distributed Cognition: Domains and Dimensions". *Pragmatics & Cognition* 14(2). 235 – 247.
Varela, Francisco J.,1987. "Laying Down a Path in Walking". In William I. Thompson (ed.) *Gaia: A Way of Knowing. Political Implications of the New Biology* 48 – 64. New York: Lindisfarne Press.
Wittgenstein, Ludwig. 2009. *Philosophical investigations* (4th ed.). Oxford: Blackwell Publishing.

Magdalena Zawisławska, Maciej Ogrodniczuk and Michał Szczyszek

Indirect Relations and Frames: Coreference in Context

Abstract: The main goal of this paper is to analyze coreference in the broader context of semantic and pragmatic relations in Polish discourse. The first section describes problems with the annotation of indirect relations in Polish using a model compiled from previous annotation projects. We focus on the creation of a typology of indirect relations within a discourse. However, while the annotation revealed many new interesting features of coreference, we noticed that further expansion of relations would hinder the process of annotation. Therefore, a different approach had to be tested. An experiment using frame semantic annotation was then conducted. Frame semantics offers a more integrated view of coreference in context—that is, other types of semantic and pragmatic concatenations in the world of discourse and the paralinguistic world. Further, using a frame analysis is a very promising method—not only does it widen coreference annotation and allows one to consider the broad context, but it also predicts the level of text complexity.

1 Introduction

Solving the problem of coreference in Polish has not been a main object of academic interest so far. The first project solely dedicated to the problem was CORE,[1] which resulted in the creation of one of the world's largest corpora of coreference (Polish Coreference Corpus),[2] as well as several important interfaces[3] and a fully-fledged automated resolution system (Ogrodniczuk *et al.* 2015). Developing an

1 A Polish National Science Centre grant 6505/B/T02/2011/40: Computer-based methods for coreference resolution in Polish texts (Pol. Komputerowe metody identyfikacji nawiązań w tekstach polskich), 2011–2014.

2 http://zil.ipipan.waw.pl/PolishCoreferenceCorpus

3 http://zil.ipipan.waw.pl/PolishCoreferenceTools

Magdalena Zawisławska, University of Warsaw (zawisla@uw.edu.pl), **Maciej Ogrodniczuk**, Institute of Computer Science, Polish Academy of Sciences (maciej.ogrodniczuk@ipipan.waw.pl), **Michał Szczyszek,** Adam Mickiewicz University in Poznań, Institute of Polish Philology (szczysze@amu.edu.pl)

https://doi.org/10.1515/9783110702286-013

annotation procedure triggered many questions concerning non-nominal coreference, borderline examples of quasi-coreference, indefinite pronouns, predicatives, more subtle semantic and pragmatic relations in a discourse, and—last but not least—the role of a paralinguistic context. Therefore, the next natural step was to broaden the description of reference by including indirect relations and creating a wider typology of referential phenomena. Although the new approach provided new insights into coreference resolution, it was still not enough to solve more elusive semantic and pragmatic relations mainly based on contextual and paralinguistic knowledge. Consequently, we conducted an experiment involving interpretative frames. We intended to learn whether frame semantics could be a useful device to resolve certain problems with context and coreference mutlifaced correlations in a discourse. In the first part of this paper, we will describe indirect relations used in an annotation of the Polish Coreference Corpus. In the second part, we will concentrate on the results of the experiment using frame semantics. We argue that this approach is not only useful in coreference resolution, but can also help to detect the level of discourse complexity.

1.1 Initial classification of indirect relations

In the context of referential relations, *direct relations* link fragments of text (mentions) referring to a common discourse world object while *indirect relations* (also called bridging or associative) link mentions referring to distinct—yet somehow contextually (semantically and/or pragmatically) related—entities or events to influence textual coherence.

The two most common examples of indirect relations come from Clark (1975):

(1) I walked into the room. The chandeliers sparkled brightly.
(2) John was murdered yesterday. The murderer got away.

Example (1) illustrates a case of nominal indirect reference by association, with an interpretation of the second nominal ('the chandeliers') that is closely dependent on a nominal in the first sentence ('the room'). Example (2) presents a case of indirect reference to an event by characterization, with the relation characterizing a role played by an agent ('the murderer') of an aforementioned event.

Several classifications of indirect relations have been proposed, both in theoretical works (Clark 1975, Prince 1981, Löbner 1996, Asher & Lascarides 1998) as well as in guidelines of former annotation projects (Poesio *et al.* 1997; Poesio 2000; Gardent *et al.* 2003; Poesio *et al.* 2004; Poesio & Artstein 2008). Clark's standard classification lists *set membership*, *indirect reference by association*

(necessary/probable/inducible parts), *indirect reference by characterization* (necessary/optional roles), *reason*, *cause*, *consequence* and *concurrence* as indirect relations. Gardent *et al.* (2003) further split the most important set membership class into *set–subset*, *set–element*, *whole–part*, *whole–piece,* and *collection–member*, and describe the potential temporal aspect of the relation (*whole–temp.-subpart*). They also introduce several thematic relations (*individual–function*, *individual–stuff*, etc.), a definitional relation (*individual–attribute etc.*) and co-participant and non-lexical relations (the latter referring to discourse structure or world knowledge). The approaches used by Poesio & Artstein (2008) as well as Recasens *et al.* (2007) limit the number of distinct relations significantly, most likely due to practical reasons of corpus annotation. Irmer (2010) splits indirect references into mereological and frame-related (thematic, causal, spatial, temporal). The Greek Coreference and Bridging Team (2014) as well as the Prague Dependency Treebank team (Zikánová *et al.* 2015: chapter 4) follow Gardent *et al.*'s approach in most respects, reserving an additional label (*rest*) to indicate a further unspecified group of *other* relations.

Based on the above survey, the classifications of indirect relations found in the literature were collated into a common typology presented in Figure 1.

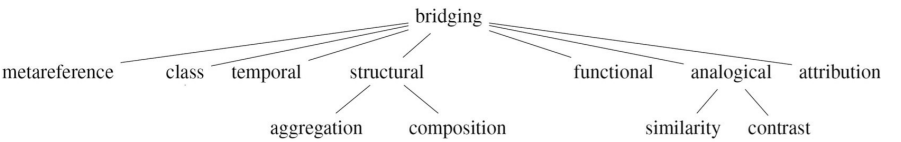

Figure 1: Literature-based initial classification of indirect relations.

The *metareference* relation is intended to model relations such as *has–model*, *has–name* or *has–label*. The *class* relation signals a class–instance relationship. The *temporal* relation is used to represent temporal aspects of the object (e.g. 'pre-war Warsaw' and 'Warsaw of today'). *Structural* or meronymic relations gather Clark's aggregation (*set–subset*, *set–element*) and composition (*whole–part*) as well as *whole–portion* or *whole–substance*.[4] The *function–object* relation is a general placeholder for various causal or thematic relations. *Analogical* relations are further split into *similarity* and *contrast* relations while *attribution* is introduced to represent relations between an object and someone's opinion on the object (i.e. what is believed, doubted etc.) or to indicate uncertainty about the nature of identity between two mentions.

4 See also (Winston *et al.* 1987) for a detailed classification of meronymic relations.

The proposed classification was initially validated in the Polish Coreference Corpus (Ogrodniczuk *et al.* 2015: chapter 8) featuring annotation of 'quasi-identity' relations,[5] i.e. relations distorting or distinguishing properties of an object, metaphorical relations between substance and container, set–element relations or other relations not characterized by identity or non-identity. Since such definition partially overlaps with the initial classification of indirect relations and since over 5100 instances of such relations were marked, quasi-identity made a useful initial resource for corpus-based validation of the typology. A randomly selected 5% of all quasi-identity relations were investigated independently by two annotators who eventually compared their decisions to produce an adjudicated result, which is presented in Table 1.

Table 1: Adjudicated annotation results with the preliminary classification of indirect relations.

Relation	Count	Relation	Count
Structural	**122**	**Metareference**	**3**
Aggregation	105	**Temporal**	**1**
Composition	17	**Other**	**10**
Class	**44**	Effect	8
Attribution	**13**	Predicate	9
Functional	**6**	Contextual	1
Analogical	**5**	**Error**	**43**
Similarity	3	Coreference	17
Contrast	2	Apposition	11
		Other	15

The error analysis showed many discrepancies between the two annotations that were attributed to annotation categories that were too broad, as well as excessively vague category definitions. Most of the relations were structural (60%), which is compatible with Gardent *et al.*'s findings.[6] At the same time, this category included extremely diverse examples that called for its division into subcategories. Many errors corresponded to incorrect classification of phenomena previously considered directly coreferential; e.g. appositions were (textually) joined to respective mentions rather than linked with any relation. A relatively small number of *other* relation subtypes shows that the annotation set can still contain interesting cases to be further investigated and introduced into the final typology.

5 Cf. Recasens *et al.* (2012) for a definition of 'near-identity' and Ogrodniczuk *et al.* (2015: chapters 1.5 and 5.1.2) for a discussion of the concept and its relation to quasi-identity.

6 See Gardent *et al.* (2003:, Figure 5).

2 Toward the Framework of Referential Relations

Two more phenomena were observed during our preliminary annotation: (i) the actual presence of relation properties (not only temporal aspect but also contextual dissimilation of properties of the referenced object), and (ii) the non-referential character of certain relations. This first finding was developed into the concept of *facets*, i.e. properties that changed the interpretation of the (base) relation or signalling its incompleteness. Figure 2 presents the facets that were identified in the process.

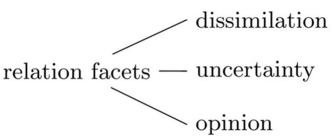

Figure 2: Classification of relation facets.

Dissimilation presents a certain feature of an object in a way that signals difficulties with deciding on its coreferentiality or non-coreferentiality with a closely related mention. The two most common cases of dissimilation are: 1) distortion of properties of the object resulting in an interpretation that refers to some 'meta-object', or 2) distinguishing a given property of an object that results in 'splitting' the object into two different references. *Uncertainty* represents the indeterminateness of a pair of objects (if expressed by the speaker) but also a conditional or propositional reference. The *opinion* facet builds on a previously used attribution that assigns subjectivity to the link, as expressed by the speaker.

The second conclusion led to the separation of relations into two sets of strictly referential and *evidence relations*, i.e. relations that signal neither direct nor indirect reference but provide information to prove or disprove the existence of a referential relation in the process of discourse analysis. Figure 3 presents the classification of supporting and excluding evidence relations building on previous classifications of indirect associations.

As compared to the initial approach, the *metareference* category now supports relations due to its conceptual rather than direct links between the expression content and the object it refers to. *Comparison* is used instead of *similarity*, and is also non-referential by nature. *Predicative expressions* are also not treated as referential in our annotation model, following Padučeva (1992: 113), because predicates tend to name attributes of the object rather than refer to it. The *other* category was introduced for cases when new types of relations could be distinguished. The *functional* relation was removed as we suspected it belongs to gen-

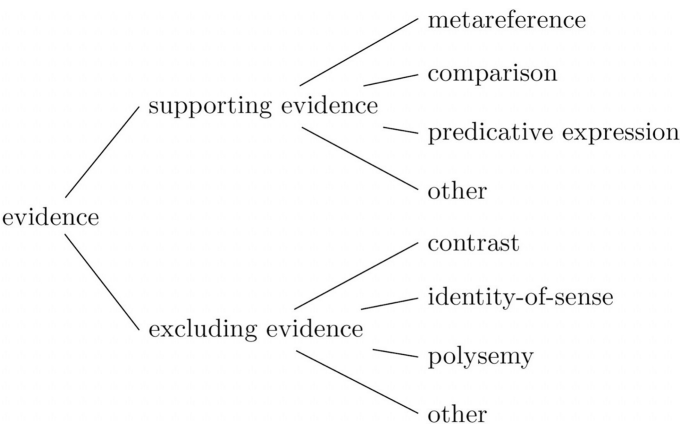

Figure 3: Classification of evidence relations.

eral discourse relations rather than supporting decoding of reference—this needs to be confirmed in the course of further annotation experiments.

Excluding (negative) relations were distinguished to signal obvious evidence of non-coreferentiality of two objects referred to in the text, such as juxtaposition of two mentions (*contrast*), relation between mentions referring to different objects 'of the same type' (*identity-of-sense*) or *polysemy*. The *other* category was also introduced.

After splitting the relation set into referential and evidence, the final classification of referential relations used in the large-scale annotation is presented in Figure 4. The *bound anaphora* category was added to the typology to signal common situations when a pronoun functions as a variable corresponding to an element of a quantified set. The distinction between *aggregation* and *class* was also removed due to its vagueness in practical contexts.

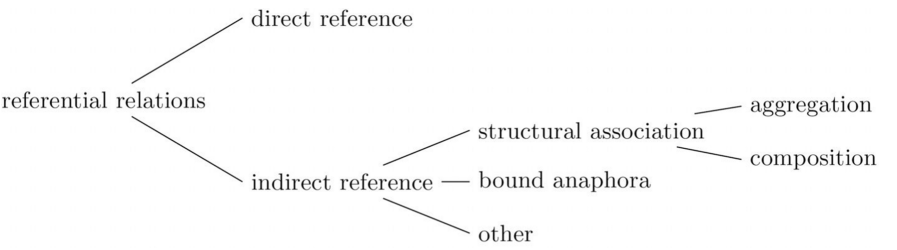

Figure 4: Classification of referential relations.

3 Annotation of Bridging Relations in Polish

After performing the initial experiments described in the previous section, a large-scale annotation of bridging relations commenced. The annotation was carried out over the complete Polish Coreference Corpus (1653 texts, 250 – 350 words each) by three annotators per text. Taking into account the practical reasons presented by other annotation projects, the proposed framework was intentionally general, with *other* classes left in the typology to be further specified by means of providing detailed comments when such choices were made. Table 2 presents statistics of identified relations averaged per relation type.

Table 2. Relation statistics (average counts per relation type).

Relation type	Relation subtype	Count
Indirect	Aggregation	7713
	Composition	2494
	Bound anaphora	312
	Other	2211
Supporting	Metareference	217
	Comparison	177
	Predicative	1945
	Other	1544
Excluding	Contrast	998
	Identity-of-sense	550
	Polysemy	182
	Negation	84
	Other	84

An interesting aspect of the annotation outcome was a relatively high amount of *other* relations (over 20 % of the total relation count). This finding called for a deeper investigation to verify whether it accounts for annotators' negligence or signals finer categories unrecognized in previous analyses. A semantic frame-based approach was used in an attempt to apply a finer-grade typology of such relations.

4 Fine-grade Typology of *Other* Relations

The *other* relations turned out to be very complicated in terms of their categorization and it was therefore necessary to develop a more detailed typology. A large set

of *other* relations, containing several thousand confirmations annotated in the corpus, can be accurately tagged by qualifying individual relationships (pre-labelled as *other*) to narrower (cognitive) subcategories. Below is a theoretical proposition of the typology of the detailed *other* relations, which was developed during the annotation of the whole corpus (*a posteriori*). It was created as a result of the analysis of the extracted relationships—first automatically, and then verified manually by human annotators. As a result, this typology has no *a priori* properties as theoretical analyses for annotators. However, after the end of the main (first) annotation cycle of the corpus, this *a posteriori* developed typology was compared with the one known from the theoretical proposals in the literature. The classification was then supplemented with any new categories provided by the *a posteriori* analysis. This improved typology was used in the second annotation cycle, which focused only on the annotation of *other* relations. This second annotation cycle resulted from the need to verify supplemented typology and was conducted on a part of randomly selected *other* relations (with hundreds of relationships of this type).

The typology of *other* relations is the following:

a) after the main (first) annotation cycle:
   - cause;
   - condition;
   - effect;
   - purpose;
   - consequences;
   - time (for convergence / identical time events);
   - succession of time (one event follows another and the two events are linked together; the distance in time can be short or long);
   - place (for convergence / space identity of events);
   - direction (understood physically, not as the culture of mainstream philosophical direction);
   - relationship (wife–husband, father in law–daughter in law, child–parent);
   - feature (object, not a function);
   - *varia*.

b) The *varia* sub-category was divided into more specific categories. Some of them are the result of the *a posteriori* typology due to a proposal that was found in the previously mentioned theoretical literature. Below are the supplements derived from the literature:
   - possessive type (e. g. your);
   - part of a whole;
   - part of something;
   - reference (e. g. one student–one computer);
   - convergence.

c) On the other hand, as a result of the second cycle annotations (only *other* relations) the above typology developed in the first cycle annotation has been completed with these additional categories:
   - opposition;
   - resources;
   - driving force;
   - topic / theme / subject (of the: conversation, debate, discussion);
   - agent;
   - parent–subordinate;
   - dependence at work;
   - method;
   - community;
   - product;
   - creation;
   - whole;
   - action;
   - agent–follower;
   - value (financial).

5 Frame-based Analysis of Indirect Relations

The complexity of the *other* relations indicates that coreference resolution needs a more sophisticated and deeper semantic description that could reflect the relatedness between a paralinguistic context and a world of discourse. As part of the COTHEC project, we decided to conduct an experiment and annotate the 21 texts from the 'long texts' subcorpus of PCC (the complete press articles from the Rzeczpospolita Corpus) using the methodology of frame semantics as a tool.

Frames are generally defined as the mental structures that organize human experiences. Fillmore (1982) posits that the meaning of lexical units, phrases, and grammatical and syntactic constructions resides in schematic phenomena, such as our beliefs, experiences or typical actions. Interpretative frames are usually evoked by lexemes: "The framing words in a text reveal the multiple ways in which the speaker or author schematizes the situation and induce the hearer to construct that environment of the text world which would motivate or explain the categorization acts expressed by the lexical choice observed in the text" (Fillmore 1982: 122).

In addition, an interpreter of a text can use much broader contextual knowledge, e. g. their cultural experience. Fillmore also notices the role of interpretative frames in text comprehension: "Interpretative frames can be introduced into

the process of understanding a text through being invoked by the interpreter or through being evoked by the text. A frame is invoked when the interpreter, in trying to make sense of a text segment, is able to assign it an interpretation by situating its content in a pattern that is known independently of the text. A frame is evoked by text if some linguistic form or pattern is conventionally associated with the frame in question. For example, the sentence 'We never open our presents until the morning' makes no mention of Christmas, yet interpreters who share certain cultural experiences, would immediately (in the terminology suggested here) invoke a Christmas context; replace the simple noun *present* with *Christmas presents* and we have introduced a word that evokes that same context" (Fillmore 1985: 232).

Ziem (2014) emphasizes the fundamental role of semantic frames in the reference. In his view, frames represent encyclopaedic knowledge that can be understood as sets of propositions—an evoked frame corresponds to the referential content of a proposition and a referent of the predication is a filler or default value of the frame's slot: "A linguistic expression refers to a cognitive unit by evoking a frame, which then opens a potential reference area [...]. Evocation of a frame corresponds to the cognitive act of referentialization. Frames—as units in the 'projected world'—serve as projection areas for referentiality. [...] In any case, referentialization is an act that creates order by co-activating the default values of the evoked frame in its very performance. Using the word *unicorn* to refer to a particular imagined object presupposes, for example, actualized assumptions about the external properties of unicorn" (Ziem 2014: 251–252).

Therefore, an interpretative frame seems to be a key factor that helps an interpreter to establish coreferential expressions in a text, due to the fact that a frame not only includes purely semantic knowledge, but also incorporates broad pragmatic and paralinguistic context. It seems that interpretative frames, in fact, represent a broad context of a world of discourse. The same role of FrameNet ontology in conference resolution is brought out by J. Zheng *et al.:* "External knowledge-bases such as FrameNet [...], Wikipedia, Yago [...] and Freebase [...], can be used to provide *global context*, and there is a strong need for coreference resolution systems to accurately use such sources for disambiguation" (Zheng *et al.* 2013: 153).

Unfortunately, The Berkeley FrameNet is not suitable for the COTHEC purposes because of the many semantic differences between Polish and English. Therefore, the annotation of the subcorpus in the COTHEC project was conducted

with a set of frames that were built from scratch,[7] e.g. PERSON, SOCIETY, ART, INDUSTRY, ARMY, CRIME, FAITH, SCHOOL, SCIENCE, TRADE, COURT, SPORT etc. The texts used in the experiment had diverse topics: the military coup d'état that took place in Greece on 21 April 1967, Saint Adalbert, Satanism in Polish schools, loyalty cards, judicial systems in different countries, consumer protection in Poland, sumo wrestling, energy production, courts in in Russia, cybernetics, textbooks, the Congress of Polish Football Association, the Andrzej Gołota vs. Michael Grant match, the Oscar Gala in 1999, etc. The average number of activated frames in the analyzed subcorpus was 10.5, while the average number of frame elements was 28.5. The described texts differed quite significantly. The texts with the smallest number of activated frames were: text 063 (about Golota's bout), with only 4 active frames, and text 084 (about the market of school books), with 5 frames. The text with the largest numbers of activated frames—18 frames—was text 029 (about General Wieniawa-Długoszowski). The results of the frame annotation are shown in Table 3. Table 4, in turn, presents the number of mentions, sets, singletons, the longest set, and an average number of mentions in one set.

Annotation using frame semantics revealed another interesting feature of the analyzed subcorpus—the level of complexity of a text. Furthermore, our experiment's results allowed us to anticipate the occurrence of certain words in a text. Below we discuss the most interesting results of the texts analyzed with frames.

6 Discussion

Example 1

In text 084 (with the dominant frame: SCHOOL), not only the sum of frames, but also the number of elements, references, sets, singletons, mentions in the longest set, and mentions in sets are below average in the subcorpus. By contrast, in text 012 (with the dominant frame: GOVERNMENT), the sum of frames is slightly above average and there are significantly more elements that are activated. The longest set of text 012 is also above average.

Both these texts look equally complex, but text 012 (GOVERNMENT) is more compact, which makes its composition better, whereas text 084 (SCHOOL) looks like it breaks into 2 or even 3 different fragments. We assume that this impression is a result of the length of the coreference chains. Text 084 (SCHOOL) has an

7 The Polish FrameNet exists (http://www.ramki.uw.edu.pl/ramki/), but it consists of only 200 verbs and is therefore excluded from this experiment.

Table 3: Frames and their elements activated in the analyzed texts.

Text number	Number of frames	Number of elements	Dominant frame(s)
084	5	19	SCHOOL
012	12	43	GOVERNMENT
043	11	36	SOCIETY, FAITH
021	11	35	SOCIETY
068	9	17	TRADE
037	9	25	COURT
127	9	27	TRADE
182	13	32	SPORT
115	9	28	INDUSTRY
095	13	26	COURT
101	10	19	MAN, MACHINE
027	8	27	OFFICE, SPORT, MAN
063	4	11	SPORT
034	6	28	ART
024	15	28	OFFICE
029	18	52	ART, MAN
022	14	27	TIME, SOCIETY, MAN
032	14	45	TRADE
057	13	29	SCIENCE
036	10	29	INDUSTRY
026	6	16	TRADE, OFFICE
AVERAGE:	10	29	

average of 3.8 mentions per set and text 012 (GOVERNMENT) has an average of 4.7 mentions per set. Additionally, in the first text the longest coreference set contains only 15 mentions, while the second one contains 64 mentions. We assume that the compactness and better composition of text 012 (GOVERNMENT) is determined by much longer coreferential sets.

Both texts have nearly average activated frames, but with a considerably different number of elements: text 012 above average, text 034 substantially lower than average. In both cases, the numbers of references and singletons are comparable, but clearly above average. However, in the case of sets, text 012 surpasses not only text 034, but also the average number of sets in the whole analyzed subcorpus.

Table 4: Statistics of mentions in sets in the analysed texts.

Text number	Number of mentions	Sets	Number of singletons	The longest set	The average number of mentions in one set
084	451	55	242	15	3.80
012	667	62	376	64	4.69
043	871	88	457	130	4.70
021	869	92	416	53	4.92
068	679	72	438	19	3.35
037	745	89	417	17	3.69
127	1002	75	641	75	4.81
182	584	54	362	30	4.11
115	441	51	225	18	4.24
095	437	37	316	13	3.27
101	441	58	263	16	3.07
027	461	36	246	54	5.97
063	383	28	185	76	7.07
034	429	47	235	26	4.13
024	416	59	222	19	3.29
029	578	64	340	55	3.72
022	492	61	244	16	4.07
032	912	92	310	78	6.54
057	490	46	285	31	4.46
036	751	85	437	23	3.69
026	525	67	271	14	3.79
AVERAGE:	601	63	330	40	4.53

Table 5. Theme variation statistics.

	Example 1		Example 2		Example 3	
Text number	084	012	021	115	063	127
Dominant frame(s)	SCHOOL	GOVERNMENT	SOCIETY	INDUSTRY	SPORT	TRADE
Frames	5	12	11	9	4	9
Elements	19	43	35	28	11	27
Mentions	451	667	869	441	383	1002
Sets	55	62	92	51	28	75
Singletons	242	376	416	225	185	641
Longest set	15	64	53	18	76	75
Average number of mentions in one set	4	5	5	4	7	5

Example 2

We compare here text 021 (SOCIETY) and text 115 (INDUSTRY). Text 115 (INDUSTRY) contains a very large amount of encyclopedic information about the judiciary system in Russia. Text 021 (SOCIETY) focuses on a much narrower topic—Satanism among young people. At the same time, there are many references to issues such as religion, the role of metal bands, and school. Such diverse themes make this text much longer and result in a bigger number of sets, and a relatively higher average number of mentions in one set (4.92). Also the longest set in text 021 considerably exceeds the average in the subcorpus. Text 115 (INDUSTRY), on the other hand, focuses solely on one topic (the judiciary system); therefore, it presents a very limited lexical resource. The average coreferential set in text 115 is apparently smaller (4.24), and the longest set contains merely 18 mentions, which is far below the average in the corpus.

Example 3

In text 063 (SPORT), the number of activated frames, elements, sets and singletons are considerably below the average in the analyzed subcorpus. Nonetheless, the longest set in the text is quite extensive as it contains 76 coreferential mentions and its median set is the largest in the whole subcorpus—it contains 7.07 mentions. However, this text is not very complicated from the perspective of its composition and semantics. In contrast, text 127 (TRADE) has a number of frames, elements and sets quite close to the average amount (the number of sets is only slightly higher). Text 127 is much more interesting and more complex than the previous case, despite having only 4.8 mentions per average set. In this comparison the most important factor is the length of the median set of coreferential mentions that indicates how many referents come out in a discourse. The unusual extent of the average set in text 063 shows that the text has a very restricted number of referents.

To sum up, the data of the entire analyzed subcorpus show the following:

- **number of activated frames:** the smallest number of active frames (4) is in text 063 (SPORT), and the biggest (18) is in text 029 (ART and MAN);
- **number of frame elements:** text 063 (SPORT) contains the smallest number of elements (11), and text 029 (ART and MAN) contains the biggest (52);
- **number of mentions:** the smallest set of mentions (383) can be found in text 063 (SPORT), and the biggest (1002) can be found in text 127 (TRADE);
- **number of coreferential sets:** text 063 (SPORT) contains the smallest number of sets (28), and text 032 (TRADE) contains the biggest number of sets (92 sets);

- **number of singletons:** the smallest set of singletons (185) can be found in text 063 (SPORT), and the largest (641) can be found in text 127 (TRADE);
- **length of sets of coreferential mentions:** text 043 (SOCIETY, FAITH) contains the longest set (130), and text 095 (COURT) contains the shortest set (13);
- **average number of mentions in one set:** text 101 (MAN, MACHINE) has the smallest average number of mentions in one set (3.07), and text 063 (SPORT) has the greatest number (7.07).

Therefore, it seems that we can determine the correlation between the text semantic and/or composition complexity depending on the distinguished main factors. A frame analysis indicates that the least complex tests are those that activate only a very narrow interpretative context (that is, frames) and contain the smallest number of other important indicators, i. e. frame elements, as well as sets, mentions, and singletons. The uncomplicated texts, however, may consist of quite a lot of mentions in one set, which signals low numbers of referents. In this case, the length of a set is inversely proportional to the complexity of a text. As the values of those most important indicators rise, texts become more and more complicated, as was shown in the analysis of the three main examples and by the summary figures for the whole corpus. This preliminary conclusion requires further studies with a larger database. The next step should concentrate on developing more formal and automatic methods that can measure the complexity of a text taking into consideration the main factors that have been pointed out in this paper. We can theoretically assume that if in a given text a number of activated frames is threefold smaller than the average for the corpus, it means that the text is monothematic, and semantically and compositionally uncomplicated. It would also be possible to predict the appearance of a particular lexicon in such a text. A more complex and sophisticated text would contain a number of elements much greater than average. In this case, the predictability of word emergence in the text is less likely due to more 'thematic fields' being activated in the discourse.

7 Conclusions

We proposed a new typology describing very fine-grained indirect relationships that is more extensive than those known so far. It is the result of work carried out *a posteriori*, and is therefore not the outcome of top-down assumptions. Furthermore, it was compared to other typologies and enriched by several other elements. As a result, we believe, our typology is able to cover all (or almost all) types of relationships that occur within an unspecified relation called *other*.

Both approaches presented in the paper allow us to include into a conference annotation elements of a broad context. The proposed classification of indirect relations demonstrates the multidimensional aspect of the coreference phenomenon, whereas the frame annotation emphasizes an integrated view of discourse.

Acknowledgements
This paper was funded by the Polish National Science Centre grant 2014/15/B/HS2/03435: Unified theory of coreference in Polish and its corpus-based verification (Pol. Ujednolicona teoria koreferencji w języku polskim i jej korpusowa weryfikacja).

References

Asher, Nicholas and Lascarides Alex. 1998. "Bridging". *Journal of Semantics* 15(1):83–113.

Clark, Herbert H. 1975. "Bridging". In *Proceedings of the 1975 Workshop on Theoretical Issues in Natural Language Processing*. TINLAP 1975. Stroudsburg, PA, USA. Association for Computational Linguistics. 169–174.

Fillmore, Charles J. 1982. "Frame semantics". In: *Linguistics in the Morning Calm*. Seoul, South Korea. Hanshin Publishing Co. 111–137.

Fillmore, Charles J. 1985. "Frames and the semantics of understanding". *Quaderni di Semantica* vol. 6.2, 222–254.

Gardent, Claire, Manuélian Hélène and Kow Eric. 2003. "Which Bridges for Bridging Definite Descriptions?" In *Proceedings of the EACL 2003 Workshop on Linguistically Interpreted Corpora*. 69–76.

GCBT: Greek Coreference & Bridging Team. 2014. *Coreference & Bridging Annotation Guidelines*. http://gdt.ilsp.gr/guidelines/coreference/coreferenceannotationguidelines.pdf.

Irmer, Matthias. 2010. *Bridging Inferences in Discourse Interpretation*. Ph.D. thesis, University of Leipzig.

Löbner, Sebastian. 1996. "Definite Associative Anaphora". In: Botley, Simon (ed.) *Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Resolution Colloquium* (DAARC 96). Lancaster University. 28–39.

Ogrodniczuk, Maciej, Głowińska Katarzyna, Kopeć Mateusz, Savary Agata and Zawisławska Magdalena. 2015. *Coreference in Polish: Annotation, Resolution and Evaluation*. Walter De Gruyter.

Padučeva, Elena Victorovna 1992. *Wypowiedź i jej odniesienie do rzeczywistości. Referencyjne aspekty znaczenia zaimków*. PWN, Warszawa.

Poesio, Massimo and Artstein Ron. 2008. "Anaphoric Annotation in the ARRAU Corpus". In: Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, and Daniel Tapias (eds.) *Proceedings of the 6th International Conference on Language Resources and Evaluation* (LREC 2008). Marrakech, Morroco. European Language Resources Association. 1170–1174.

Poesio, Massimo, Vieira Renata and Teufel Simone. 1997. "Resolving Bridging References in Unrestricted Text". In: *Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*. 1–6. https://www.aclweb.org/anthology/W97-1300.

Poesio, Massimo, Delmonte Rodolfo, Bristot Antonello, Chiran Luminita and Tonelli Sara. 2004. *The VENEX Corpus of Anaphora and Deixis in Spoken and Written Italian*. Unpublished manuscript. https://www.academia.edu/download/48751723/VENEX04.pdf

Poesio, Massimo. 2000. *The GNOME Annotation Scheme Manual*. Technical report, University of Essex, United Kingdom.

Prince, Ellen F. 1981. "Toward a taxonomy of given-new information". In: Phillip Cole (ed.) *Syntax and semantics: Vol. 14. Radical Pragmatics* Academic Press, New York. 223–255.

Recasens, Marta, Martí Antonia and Taulé Mariona. 2007. "Where Anaphora and Coreference Meet. Annotation in the Spanish CESS-ECE Corpus". In: *Proceedings of RANLP* 2007. Borovets. Bulgaria. 504–509.

Recasens, Marta, Martí Antonia and Orasan Constantin. 2012. "Annotating Near-Identity from Coreference Disagreements". In: Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Dogan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis. (eds.) *Proceedings of the 8th International Conference on Language Resources and Evaluation*. (LREC 2012). Istanbul, Turkey. European Language Resources Association (ELRA). https://www.aclweb.org/anthology/L12-1391/.

Winston, Morton E., Chaffin Roger and Herrman Douglas. 1987. "A Taxonomy of Part–Whole Relations". *Cognitive Science*. 11(4): 417–444.

Zheng, Jiaping, Vilnis Luke, Singh Sameer, Choi Jinho D. and McCallum Andrew. 2013. *Dynamic Knowledge-Base Alignment for Coreference Resolution*. In: Proceedings of the 17th Conference on Computational Natural Language Learning. The Association for Computational Linguistics. 153–162.

Ziem, Alexander. 2014. *Frames of understanding in text and discourse: Theoretical foundations and descriptive applications*. Amsterdam: John Benjamins Publishing.

Zikánová, Šárka, Hajičová Eva, Hladká Barbora, Jínová Pavlina, Mírovský Jiří, Nedoluzhko Anna, Poláková Lucie, Rysová Kateřina, Rysová Magdaléna and Václ Jan. 2015. *Discourse and Coherence. From the Sentence Structure to Relations in Text*. Prague: Institute of Formal and Applied Linguistics, Charles University.

Author's Index

Subject Index